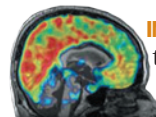


THIS WEEK

EDITORIALS

FUTURE SHOCKS The enduring lure of time travel and other fictions **p.594**

WORLD VIEW The United States heads in the right direction on science **p.595**



IMAGING Scans to find the tangled tau proteins in live brains **p.597**

All together now

Proposals to bring hydrofluorocarbons under the auspices of the Montreal Protocol provide a simple test of the international community's commitment to tackling climate change.

Four years ago, this journal endorsed a simple idea: use the world's most successful international environmental treaty, the Montreal Protocol on Substances that Deplete the Ozone Layer, to solve the world's most difficult environmental problem — global warming. Two years ago, *Nature* chided a handful of countries for blocking the path forwards, chiefly China, India and Brazil. Today, we are left with one major holdout: India.

As the latest negotiations over the future of the Montreal Protocol wrapped up in Bangkok on 25 October, India found itself increasingly isolated, and rightly so. On the table were a pair of amendments that would pull the regulation of hydrofluorocarbons (HFCs) out of the United Nations climate framework and into the Montreal Protocol's portfolio. HFCs replaced the infamous ozone-eating chlorofluorocarbons (CFCs), and were successfully created and deployed under the protocol to protect the ozone layer, which is now on the mend. However, they are powerful greenhouse gases. The simplest solution is to use the same tool that bred HFCs to phase them out. India took the lead in blocking consideration of the amendments.

In all likelihood, their acceptance is just a matter of time. Most countries have long supported the idea, and early objectors such as Brazil and South Africa have come around. US President Barack Obama brought Chinese President Xi Jinping on board in June, and the leaders of the Group of 20 (G20) nations — including India — gave their endorsement in September. The fact that India is on the losing side of this debate makes its renewed intransigence all the more galling. But there is hope: after the September G20 meeting, Obama and Indian Prime Minister Manmohan Singh agreed to launch negotiations over the issue.

Without action, HFC usage will rise sharply owing to more demand for refrigeration and air conditioning in the developing world. The Intergovernmental Panel on Climate Change's recent assessment of the science underlying global warming highlighted the challenges ahead, and its assessments of adaptation and mitigation measures will soon follow. The message is clear: delaying action will increase the severity of climate impacts, and India is well aware that these are likely to hit hardest in developing countries. Another year of haggling will probably not change the outcome of negotiations, but it will cost the world precious time.

Meanwhile, many in industry are gearing up to replace HFCs with chemicals such as HFO-1234yf, jointly developed by the chemical giants Dupont in Wilmington, Delaware, and Honeywell in Morris Township, New Jersey. Created in response to European Union regulations to limit the climate impact of vehicle refrigerants, the chemical is some 325 times less powerful as a greenhouse gas than the current industry standard.

More work is needed to replace HFCs in other applications, but the Montreal Protocol's job is to harness this work and accelerate the change. One concern in the debate is how much it will cost to shift industry towards climate-friendly chemicals, and who will pay. The treaty has a well-trodden pathway: developed countries pioneer workable solutions and then help developing countries with the transition.

Another concern is little more than a turf war between the Montreal Protocol and the UN climate convention, which has jurisdiction over greenhouse gases. The proposed amendments would address this by shifting management of the chemicals into the Montreal Protocol while leaving the accounting to the convention.

Delegates at the Bangkok meeting called for a technical report on HFC alternatives and scheduled a workshop on the issue for next year. That leaves the door open to a decision before the next headline climate summit in Paris in 2015. Success on this front might restore confidence in the multilateral process and build

some much-needed momentum going into the talks. It has been clear for some time that it will be difficult, if not impossible, for the world to achieve a singular solution through the UN climate negotiations. Climate mitigation has become decentralized, and countries must use all of the tools at their disposal to reduce emissions. With a little success, diplomats may find it easier to increase political ambitions and fold these efforts into a viable climate treaty.

At stake in the Montreal Protocol talks is not just the future of one treaty, but also our legitimately shaken confidence in multilateralism. If the world cannot agree on something as simple as this, what hope is there of meaningful cooperation on the difficult issues that lie ahead? ■

"At stake is our legitimately shaken confidence in multilateralism."

Time to talk

Online discussion is an essential aspect of the post-publication review of findings.

Scientists are an opinionated bunch. From the cutting criticisms they make during peer review to bold questioning after a conference presentation, the rough and tumble of academic debate is seen as a crucial part of scientific progress. So where is the online equivalent for published papers?

To be sure, there are lively debates on blogs and social media about the merits of published work. Individual communities have formed their own central areas for engaged and informed criticism of peer-reviewed results. For example, Haldane's Sieve, a site for evolutionary geneticists, is always busy and encourages authors to write short explanations of preprint abstracts. But click on the published homes of many of these papers — the websites of the journals in which they appear — and you will find digital tumbleweed.

In recent years, authors and readers have been able to post online comments about *Nature* papers on our site. Few bother. At the Public

Library of Science, where the commenting system is more successful, only 10% of papers have comments, and most of those have only one.

Last week, the US National Center for Biotechnology Information became the latest publisher to attempt to corral this online discussion. It is opening up PubMed, its online repository of abstracts of some 23 million scientific papers, to comments. If online commenting on research papers or abstracts can flourish anywhere, surely it will be at PubMed, the site at which so many biomedical researchers begin their investigation of the literature. Visibility should be high, after the site emerges from its restricted pilot stage. And next week will see the launch of bioRxiv, the latest attempt to mimic the physics preprint site arXiv for biomedical sciences; this site will also encourage comments.

The early signs for the PubMed trial are promising. Among the first 200 comments, there has been some apparently useful self-criticism: for example, Andrew Kniss of the University of Wyoming in Laramie, posted on his own 2006 paper, saying that “our conclusions with respect to field management of the disease went beyond what the limited data could support”.

What PubMed has going for it is high traffic. But what has yet to emerge, and what could dictate how scientists use it, are the boundaries of online conduct that it will set, and how it will tune into the constructive criticism needed for genuine post-publication review of findings amid the noise and static that passes for much online debate.

Alarmed at the tenor of criticism when it concerns misconduct, some editors have tried to rein in online discussion, and to bring it within the limits of conventional debate. But attempts to dictate terms are likely to backfire. In a recent editorial in the journal *ACS Nano*, for example, the editors asserted that “the numbers of blogs, twitter messages, etc. in

which individuals accuse others of academic fraud are steadily rising” — although they did not provide evidence for this. And they asked that suspicions of plagiarism or data manipulation be reported directly to a journal, rather than posted openly online (W. J. Parak *et al.* *ACS Nano* 7, 8313–8316; 2013). It was others’ “privilege” to be able to comment on a journal’s decision on a blog afterwards, the editorial added.

Although written with concern for the fair treatment of scientists who suffer damage to their reputation when comments are made irresponsibly, the editorial raised the hackles of chemistry bloggers who have

“It makes sense to digitally tie the research and the comments together.”

pointed out egregious examples of image manipulation in papers — and who understandably consider that it is they, as much as the journals, who are doing the community the service (see go.nature.com/cplnfd). It is better to ask that debate be civil, responsible and courteous, than that it not appear online at all.

Does it matter just where online these discussions take place? Not in the short term — even the most obscure blog discussion can catch a wave and throw an academic debate into the mainstream. But if a longer-term goal is to leave some permanent signpost to help others navigate the scientific landscape, then it does make sense to digitally tie the research and the comments together.

For this to happen, PubMed Commons and digital publishers either need to become a hub for separate online discussions, or to generate community engagement directly at the sites of research papers or abstracts — something that has so far been lacking. Ultimately, the success of this worthy effort is incumbent on us all: so do visit and comment on your paper and others. Help the experiment to work. ■

Playful paradoxes

A half-century of Doctor Who has shown the dramatic possibilities of science in the arts.

November sees perhaps the most significant golden anniversary to have been celebrated anywhere or anywhere in the entirety of time and space. Yes — it will have been 50 years since the broadcast of the first episode of *Doctor Who*, the adventures of a character who has become television’s most celebrated time traveller. On page 620 of this issue, cosmologist Andrew Jaffe looks at the facts and fictions of time travel as an enduring trope in fiction. As he notes, *Doctor Who*’s generally gleeful disregard of time-travel paradoxes is all for the good. There is a danger of taking such things much too seriously. What is more important is the story itself and the situations it offers its participants.

In other words, time travel might well be impossible in real life, but so what? Playing with its paradoxes is fun and inspires millions who might not otherwise have done so to consider its possibilities — dramatic, if not physical. There is, nonetheless, a valid point to be made: if time travel is unbelievable, isn’t there a danger that readers or viewers cannot suspend disbelief for the duration? Well, yes, but if the writing is good enough, the audience can be carried along for the ride as long as it lasts, never questioning the reality of dragons or dilithium crystals, nor noting the delicious irony of a starship engineer in *Star Trek* who complains that “ye cannae change the laws of physics, Captain”. This is why authors, artists and film-makers try their best to get right small details most likely to pull audiences up rather than worrying about the overarching conceit of (say) time travel. As zoologist Adam Summers recalled when he was brought in by Pixar to consult on their animation *Finding Nemo* (2003), artists pay great attention to detail to make their acts of world creation believable, even if fish don’t actually talk (*Nature* 427, 672–673; 2004).

Some authors, however, balk at bringing unbelievable elements into their stories, especially if they see themselves as having regard

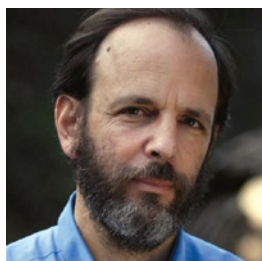
for science. In the 1980 novel *Timescape*, real-life physicist Gregory Benford grants himself the allowance for information to be transmitted backwards through time, if not the transfer of people or machinery — a limitation that he exploits to perfection in the pay-off. People in a future world doomed by ecological catastrophe remain doomed, even though they transmit their warnings so that a past world can save itself.

Another staple of science fiction is the ability of spacecraft to travel enormous distances faster than the speed of light — something else that is almost certainly impossible. In a note to his 1986 novel *The Songs of Distant Earth*, the late Arthur C. Clarke writes that it “now seems almost certain that in the real universe we may never exceed the velocity of light. Even the very closest star systems will always be decades or centuries apart: no Warp Six will ever get you from one episode to another in time for next week’s instalment. The great Producer in the Sky did not arrange his program planning that way.” Clarke uses this limitation to poignant dramatic effect; not that it prevented him from exploiting such improbabilities as tapping vacuum energy or long-term suspended animation.

Perhaps the most surprising critic of such technological fixes was the great hobbitmonger himself, J. R. R. Tolkien, as revealed in his unfinished story *The Notion Club Papers* (published posthumously in 1992 in *Sauron Defeated*, edited by his son and literary executor Christopher Tolkien). Much of the story is a discussion between academics and writers on the dishonesty of using scientific-sounding MacGuffins to get one from here to there. If one insists on doing such a thing, one might as well dream oneself to Mars or wave a wizardly wand. The story centres on criticism of H. G. Wells’s *The First Men in the Moon* (1901), in which the protagonist, one Dr Cavor, invents a material, cavorite, that provides insulation against gravity. “Gravity can’t be treated like that,” complains one of Tolkien’s characters. “It’s fundamental. It’s a statement by the Universe of where you are in the Universe, and the Universe can’t be tricked by a surname with *ite* stuck on the end, nor by any such abracadabra.” Which

suggests, perhaps surprisingly, that even if time travel and warp drives are impossible, the world’s best-selling fantasy author knew a thing or two about the general theory of relativity. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Science's rightful place is in service of society

Science policy must concentrate less on how much money is spent, and more on how to translate investments into public good, says **Daniel Sarewitz**.

Amid the mess of US politics — a pointless government shut-down, across-the-board cuts, endless partisan squabbling — now is a good moment to take stock of the fate of publicly funded science. After all, five years ago next week Barack Obama was first elected president, promising that he would “restore science to its rightful place” in US society. How has he done?

Pretty well — and the ongoing budget crisis might be the most important reason. When there is no new money to throw at science, the only way to improve its social value is to tighten how the old money is spent. And science policies under Obama are beginning to add up to a strategy to correct the greatest weakness of the US research enterprise: the isolation of the conduct of science from its use in society.

In biomedicine, the doubling of the US National Institutes of Health (NIH) budget between 1998 and 2003 did not reduce the stunningly high failure rates and costs of drug development. To confront this problem, the Obama administration created the National Center for Advancing Translational Sciences (NCATS), which was approved by Congress in December 2011. Central to NCATS' vision, says NIH director Francis Collins, are partnerships between “government, academia, philanthropy, patient advocates, and biotechnology and pharmaceutical companies to overcome translational roadblocks and offer solutions to detect, treat and prevent disease”.

Despite forecasts of doom, basic science in the United States stands preeminent, as shown by the ongoing harvest of Nobel prizes. But where is the pay-off for the rest of society? The bankruptcy of Detroit in Michigan, once the world auto-industry capital, underscores the need for new science-based technology sectors to create jobs for millions of people, yet it also makes apparent the lack of connection between scientific excellence and economic well-being. To help close this gap, the Obama administration last year created the National Additive Manufacturing Institute. Focused on three-dimensional printing, it is located in the ‘rust belt’ city of Youngstown, Ohio, and was launched with a US\$30-million government contribution matched by corporate funds. In May, the president announced three more manufacturing institutes, each to be “a regional hub designed to bridge the gap between basic research and product development, bringing together companies, universities and community colleges, and federal agencies to co-invest in technology areas”.

In climate change, more than 20 years of research have yielded little in terms of a strategy to adapt to climate impacts or mitigate their causes. The administration is wrapping up what is only the third national assessment of climate since 1990, and is calling for a new process of “sustained assessment”. National

assessments are seen as obligatory end-of-pipe summaries of knowledge. Sustained assessment is different: a continual process, according to the US Global Change Research Program, of engaging “diverse viewpoints of private industry, state and local governments, non-governmental organizations, professional societies and impacted communities” that helps scientists to “understand what information society wants and needs” and “provides mechanisms for researchers to receive ongoing feedback on the utility of the tools and data they provide”.

The Department of Energy (DOE) has long been viewed as an underperformer, beset by programmatic fiefdoms and high-profile failures. Even before President Obama was elected, Congress had authorized the creation of the Advanced Research Projects Agency-Energy to pursue high-risk, high-reward projects outside the DOE bureaucracy, but the

agency did not get funded until early in the Obama administration. A year later, the DOE launched the Energy Innovation Hubs to address challenges such as energy storage, through collaborative teams of “top talent across the full spectrum of R&D performers — including universities, private industry, non-profits, and government laboratories — integrating expertise in multiple scientific disciplines, engineering fields, and technology areas”. And last July, the department said that it was bringing basic research and energy technology research into the same administrative home to enhance “the ability to closely integrate and move quickly among basic science, applied research, technology demonstration, and deployment” — a change that should have been made 25 years ago.

What ties these initiatives together? It is the recognition that when scientists and knowledge users understand one another's evolving capabilities

and needs, resources can be allocated more effectively, and knowledge can be tested for reliability and used more efficiently. Each of the initiatives aims to foster close and persistent links between scientists and those who might benefit from scientific knowledge.

These programmes are not panaceas, and several of them have been controversial. But they move the goals of science policy in the right direction — away from an obsession with how much money is spent on science, and towards a consideration of how best to ensure that science investments turn into public value. The ‘rightful place’ of science must be created through complex institutional arrangements that allow the progress and contributions of science to emerge from its engagement with society. ■

Daniel Sarewitz is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC.
e-mail: daniel.sarewitz@asu.edu

**BASIC SCIENCE IN THE
UNITED STATES STILL
STANDS
PREEMINENT
BUT WHERE IS THE
PAY-OFF FOR THE
REST OF
SOCIETY?**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/fagcs7

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

REMOTE SENSING

Illicit gold rush in Peruvian Amazon

Gold mines across the Peruvian Amazon increased by more than 400% from 1999 to 2012, and are now the main cause of deforestation there.

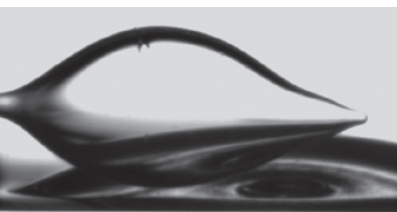
A team led by Gregory Asner of the Carnegie Institution for Science in Stanford, California, validated the results from satellite data using field sampling and high-resolution aerial surveys. These estimates are twice that of previous studies, and suggest that mining covered fewer than 10,000 hectares in 1999 but had spread beyond 50,000 hectares by September 2012.

The rate of forest loss more than tripled as gold prices rose in 2008, and was driven by small, illegal mining operations that now account for most activity in the region. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1318271110> (2013)

MATERIALS SCIENCE

Water sculptures crafted in oil

Water droplets suspended in oil can be shaped into ellipsoids, tubes or even fish-like forms (**pictured**). Thomas Russell at the University of Massachusetts in Amherst and his team added chemically modified polymers and polystyrene nanoparticles to water droplets in oil.



REBECCA BLIEGE BIRD

ECOLOGY

Hunting leads to a leap in lizards

Traditional hunting seems to boost lizard populations in Australia.

Rebecca Bliege Bird at Stanford University, California, and her colleagues found that numbers of sand monitor lizards (*Varanus gouldii*) in the Western Desert were largest where there was most hunting — lizard burrows were present in 13% of hunted plots but in only 7% of plots with least hunting. Aboriginal hunters burn small patches of land, thus

promoting habitat that favours *V. gouldii* and apparently outweighing the loss of individual lizards to Aboriginal dinner plates.

The authors suggest that extinctions of native species in the Australian desert might be linked to a decline in traditional hunters, whose cultural 'Dreaming' knowledge stresses the importance of human-made fires to the health of the landscape.

Proc. R. Soc. B 280, 20132297 (2013)

CANCER

Antibody creeps up on cancer

An antibody that becomes active only when it encounters tumours could provide a path to safer cancer drugs.

Although gentler than other cancer drugs, side effects from antibodies still limit the dosage that patients can receive. To reduce toxicity, Henry Lowman of CytomX, a biotechnology firm in South San Francisco, California, and his colleagues

designed a covert antibody — a protein-based molecule that cannot bind to its target until it faces protein-cleaving enzymes that are often abundant near tumours.

In mice, a covert antibody against a cancer-promoting protein called epidermal growth factor receptor showed little activity in the blood, yet fought tumours at levels similar to the cancer drug cetuximab. The antibody was also less toxic than cetuximab in monkeys. *Sci. Transl. Med.* 5, 207ra144 (2013)

SCIENCE/AAAS

CLIMATE SCIENCE

Meandering winds precede heat spells

Although the deadly heatwaves that afflict the United States arise from chaotic forces, they could be predicted up to three weeks before they hit.

As yet, heatwaves cannot be foreseen until about ten days in advance. However, in a 12,000-year computer simulation of general atmospheric circulation, Haiyan Teng at the National Center for Atmospheric Research in Boulder, Colorado, and her colleagues observed patterns in the meanderings of high-altitude winds that tended to precede North American heatwaves by 15–20 days.

In the model, large-scale circulation anomalies that increased the risk of heatwaves were not linked to other commonly used indicators, such as monsoons or sea surface temperatures. Monitoring mid-latitude atmospheric dynamics could allow more accurate and longer-term forecasts, the authors say.

Nature Geosci. <http://doi.org/pnt> (2013)

For a longer story on this research, see go.nature.com/fjsuha

PLANETARY SCIENCE

Why Martian craters are flat

When meteorites pummelled a young Mars, they may have obliterated enough of its crust to allow the planet's mantle to well up and trigger volcano-like activity across the red planet.

A team led by Christopher Edwards of the California Institute of Technology in Pasadena looked at 2,800 ancient Martian craters with flat floors, thought to be made of sediment layers. But data from orbiting satellites showed that materials in the crater were too rocky for this to be true. Other observations, such as the materials' mineral composition, were also incompatible with sediment.

Instead, the authors propose, meteorite impacts that fractured the planet's crust provided paths for molten rock to flow up and flood the crater floors. At a time when Earth's early oceans supported life, the surface of Mars may have been marked by magma.

Icarus <http://doi.org/pd2> (2013)

NEUROSCIENCE

Rats propelled by electric pulses

Deep-brain stimulation can help rats with injured spinal cords to walk. After severing a large portion of neural fibres in rats' spines, researchers led by Lukas Bachmann at the University of Zurich in Switzerland applied electrical pulses to a brainstem region that controls locomotion.

This allowed rats with only about 20% of intact neural fibres to walk almost normally, even though similar injuries in rats and humans cause severe impairments in walking. Even rats with less than 10% of these fibres and completely paralysed hindlimbs regained some movement when swimming.

Although more work is necessary to evaluate the approach's therapeutic potential, tapping into a person's existing motor command circuitry in the brain may provide more control over voluntary movements than do other approaches that stimulate the spinal cord itself.

Sci. Transl. Med. 5, 208ra146 (2013)

EVOLUTION

Human ancestor had small thumbs

Fossil analysis reveals that an ancestor of modern humans would have made a terrible hitch-hiker.

Past reconstructions of the hands of the hominin *Australopithecus afarensis* assigned scattered bones to individuals and single fingers.

COMMUNITY CHOICE

The most viewed papers in science

ALZHEIMER'S DISEASE

Tau tangles exposed

HIGHLY READ
on www.cell.com
in October

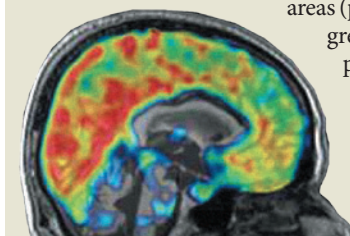
Tangles of the tau protein are found in the brains of people who have died with Alzheimer's disease, but establishing techniques to see the protein in living people has been difficult.

A team led by Makoto Higuchi at the National Institute of Radiological Sciences in Chiba, Japan, developed a class of radioactive molecule that can be used to image live brains.

The researchers first confirmed that the molecules bound to tau using brain slices from patients with Alzheimer's disease and in mutant mice with tau deposits in their brains. The team then injected the imaging molecules into human participants. In three people with probable disease, imaging using positron emission tomography showed the molecules binding in brain

areas (pictured; increasing intensity, green to red) associated with the progression of Alzheimer's; only minimal binding was observed in three healthy individuals.

Neuron 79, 1094–1108 (2013)



Campbell Rolian at the University of Calgary, Canada, and Adam Gordon at the State University of New York at Albany reanalysed an assembly of *A. afarensis* bones to better account for uncertainties in the fossil record.

Their results suggest that the hominin had shorter thumbs than was thought, with proportions more closely resembling those of gorillas. Although *A. afarensis* may have been able to bring the tips of its fingers and thumbs together, its thumbs were not long enough for the precision grip that later hominins used to craft stone tools.

Am. J. Phys. Anthropol. 152, 393–406 (2013)

called proviruses embed themselves in the DNA of immune-system cells, thwarting efforts to eliminate HIV infection. However, in laboratory tests, less than 1% of proviruses could be reactivated — the rest were thought to be so mutated that they are unable to awaken and spread.

Now, Robert Siliciano of Johns Hopkins University in Baltimore, Maryland, and his colleagues show that in people who have been treated for HIV, the proportion of potentially active proviruses is higher than assumed. The researchers sequenced DNA of 213 inactive proviruses from people with HIV and found that around 12% of proviruses still had intact genomes. That makes the reservoir of latent virus some 60 times larger than previous estimates.

Cell 155, 540–551 (2013)

VIROLOGY

Not all dormant HIV is defective

Viral 'sleepers' in human genomes could make HIV even harder to cure than expected.

Dormant HIV sequences

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

FUNDING

Horizon 2020 funds

European Union (EU) member states should not use the region's central funding to prop up national research budgets, says Anne Glover, chief scientific adviser to the president of the European Commission. Speaking in London on 22 October, Glover warned that the 28 EU member states — 23 of which have cut research spending in recent years, she said — should not expect funds from the Horizon 2020 initiative to bail out their research operations. Between 2014 and 2020, the EU's Horizon 2020 programme is set to distribute some €70 billion (US\$97 billion) for research — such as transnational projects — that would not otherwise be funded by individual states.

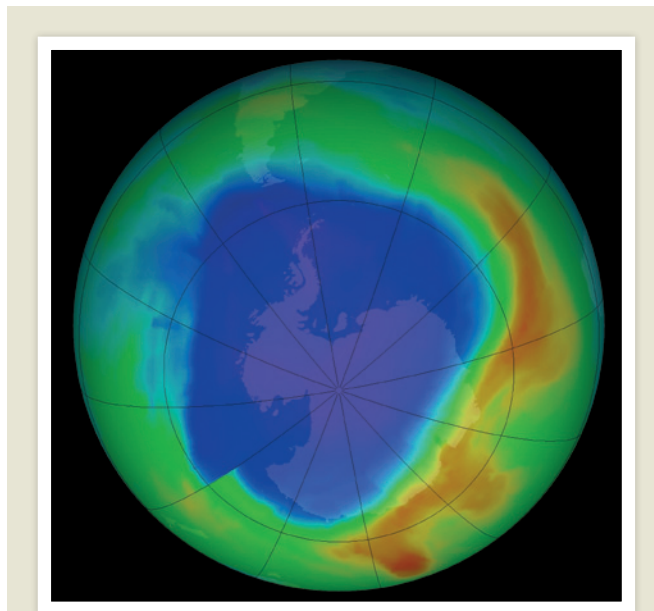
POLICY

Fishing vote

The campaign to reform Europe's fisheries inched forwards on 23 October, when the European Parliament voted against using part of a roughly €6.4-billion (US\$8.8-billion) subsidy package to fund new fishing vessels. Part of the fund will, however, support upgrades to boat engines, which some scientists say will exacerbate overfishing in European waters. Lawmakers also voted to use a portion of the money for data collection (see *go.nature.com/brbctr* and *Nature* 502, 420; 2013).

Indian power

A controversial nuclear power plant in India began generating electricity last week, after weathering years of protest (see *Nature* <http://doi.org/ckcr86>; 2011). The first unit of the Kudankulam



Ozone hole shrinks

The seasonal ozone hole over Antarctica (pictured, blue) was slightly smaller this year than in past years, measuring around 21 million square kilometres in late September, the US National Oceanic and Atmospheric Administration (NOAA) reported last week. "We cannot say that this represents recovery, but it is certainly good news," says NOAA researcher Bryan Johnson. Chlorine levels at the poles, once raised by ozone-depleting foams, sprays and refrigerants, are now on the decline, the agency notes. The use of hydrofluorocarbons (HFCs) as an alternative to these chemicals has raised other concerns, however, because HFCs are powerful greenhouse gases. Last week, negotiations in Bangkok fell short of an agreement to regulate HFCs under the Montreal Protocol on Substances that Deplete the Ozone Layer. See page 593 for more.

Nuclear Power Project in Tamil Nadu was hooked up to the country's electrical grid on 22 October. The unit will increase India's capacity for nuclear-derived electricity by 1,000 megawatts, bringing it to a total of 5,780 megawatts.

Greenland uranium

The parliament of Greenland voted 15–14 on 24 October to lift a ban on the mining of radioactive materials such as uranium, as well as rare-earth metals used in the aerospace, electronics and

energy industries. Greenland, which is a semi-autonomous territory of Denmark, has significant uranium deposits, and the policy shift could lead to significant investment from China.

Citizens weigh in

Two European campaigns last week each collected the 1 million signatures they needed to reopen political debate on controversial research topics. The Stop Vivisection campaign seeks to ban the use of animals in

research, while the One of Us campaign aims to block European Union (EU) funding for research involving human embryonic stem cells. The campaigns are among the first to succeed in the EU's new initiative in participatory democracy. Once the signatures are verified, the groups will present their cases in public hearings before the European Parliament.

Food czars faulted

Too many members of the European Food Safety Authority (EFSA) maintain close ties with the food industries they regulate, according to a report published last week by the Corporate Europe Observatory (CEO). Based in Brussels, the CEO seeks to expose the effects of lobbying on European policy-making. The report found that nearly 60% of EFSA scientific panel members had at least one self-reported conflict of interest. The EFSA's panel on dietetic products, nutrition and allergies was the worst offender, with 113 conflicts of interest found among 17 of its 20 members.

RESEARCH

Data dash

NASA has set a new speed record for beaming information between the Moon and Earth, the agency announced on 22 October. The Lunar Atmosphere and Dust Environment Explorer (LADEE), launched in September and currently orbiting the Moon, sent data encoded in a pulsed laser beam across roughly 384,400 kilometres at a rate of 622 megabits per second — five times the previous rate. The information was received by a ground station in Las Cruces, New Mexico.

GODDARD SPACE FLIGHT CENTER

LLNL LADEE is carrying NASA's first two-way communication system to use lasers instead of radio waves (see *Nature* **499**, 266–267; 2013).

PubMed comments

Got a bone to pick?

Researchers will soon be able to comment on each other's work on PubMed, an online database of life-sciences papers. The US National Center for Biotechnology Information, which runs PubMed, launched the pilot phase of its new programme, PubMed Commons, on 22 October. Although only invited participants can view and post comments at present, the programme will soon allow all PubMed authors to write comments, which will be visible to the general public, under their real names. See page 593 and go.nature.com/dgftwk for more.

TB assessment

Some 3 million people infected with tuberculosis (TB) went undetected by global health systems in 2012, the World Health Organization (WHO) said on 23 October. In its annual report on the disease, the WHO also flagged major shortcomings in the testing and treatment of people with drug-resistant forms of TB. Overall, however, the organization found that the world is on track to meet the

United Nations Millennium Development Goals of cutting TB incidence rates by 2015, and halving mortality rates compared to 1990 levels. See go.nature.com/gtuvpv for more.

Planck plug pulled

The European Space Agency gave the signal on 23 October to shut down the Planck satellite, a mission to study the magnetic afterglow of the Big Bang. Launched in 2009, Planck has created ultra-precise maps of the cosmic microwave background, helping to refine theories about the Universe's early structure and its evolution over billions of years (see *Nature* **495**, 417–418; 2013 and go.nature.com/szagdh for more.)

PEOPLE

Caltech president

Physicist Thomas Rosenbaum has been appointed as the next president of the California Institute of Technology (Caltech) in Pasadena. Rosenbaum has been provost of the University of Chicago in Illinois since 2007. There, he oversaw the establishment of the Institute for Molecular Engineering, which launched the university's first graduate programme in engineering in May 2013. Rosenbaum will take the helm at Caltech in July 2014.



Weapons lab head

Penrose 'Parney' Albright (pictured) is stepping down as director of the Lawrence Livermore National Laboratory in California (LLNL). The physicist, who has led the laboratory since December 2011, announced his resignation on 24 October, saying that he will continue to pursue his interests in US national security. Bret Knapp, who directs weapons programmes at Los Alamos National Laboratory in New Mexico, will take over on 1 November as interim LLNL director until Albright's successor is appointed.

BUSINESS

Genetics spinout

Investors hope to squeeze more profits out of the once-bankrupt firm deCODE Genetics, with the 23 October launch of NextCODE Health, a spinout company. NextCODE,

COMING UP

1–5 NOVEMBER

Genetic liver disease, liver transplants and treatments for hepatitis C are highlighted at the 64th Annual Meeting of the American Association for the Study of Liver Diseases in Washington DC. go.nature.com/y2oad

4–7 NOVEMBER

At the International Conference on Regional Climate in Brussels, topics include the use of regional climate data in disaster-risk reduction, food security and health. go.nature.com/66dlqh

based in Cambridge, Massachusetts, has secured US\$15 million in private funding to look for disease-causing variants in DNA sequence data using deCODE's database of genetic and medical records. Reykjavik-based deCODE filed for bankruptcy in 2009, and was later acquired by biotechnology company Amgen in Thousand Oaks, California (see go.nature.com/kygrxz).

Alzheimer's drug

The US Food and Drug Administration on 25 October approved the second radioactive drug to aid the evaluation of Alzheimer's disease. Vizamyl (flutemetamol) from GE Healthcare, headquartered in Pollards Wood, UK, binds to β -amyloid protein, which accumulates in the brains of people with the disease, and allows the protein to be seen by positron emission tomography imaging. A positive result should be used with other tests to diagnose Alzheimer's, the agency noted.

➔ NATURE.COM

For daily news updates see: www.nature.com/news

TREND WATCH

Reports that government researchers in Canada experience political interference in the practice and communication of science have been backed by the results of a survey of more than 4,000 scientists. The study, conducted in June and released this month, was commissioned by the Professional Institute of the Public Service of Canada, which represents workers in 40 government agencies. Since 2007, federal environmental scientists have had to get approval before speaking to the press.

STIFLING SCIENCE

A survey puts numbers to perceptions that government scientists in Canada are being muzzled.



90%
Feel that they cannot speak freely to the media about their work



48%
Had seen information withheld, causing the public or government to be misled or misinformed



86%
Could not report actions that might harm the public without fear of censure



43%
Had been asked to exclude or alter information in government documents for non-scientific reasons



50%
Had seen public health and safety compromised by political interference in science



37%
Had been blocked from answering media requests in the past 5 years

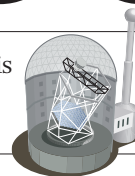
NEWS IN FOCUS

ASTROPHYSICS Study challenges Hawking theory, suggesting black holes never die **p.603**

METEOROLOGY Lightning-detection system opens up new front in forecasting **p.604**

AGRICULTURE Soil-analysis system aims to boost harvests in Africa **p.607**

ASTRONOMY South Africa's telescope tries to beat its teething problems **p.608**



ABIR SULTAN/EPA/ALAMY



A child receives an oral vaccine against poliovirus in Israel, where the virus has been silently spreading since February without causing paralytic disease.

INFECTIOUS DISEASE

Polio risk looms over Europe

Cases in Syria highlight vulnerability of nearby countries to the viral disease.

BY DECLAN BUTLER

To many Europeans, poliomyelitis is an ancient foe. But for the first time in years, there is a risk that the crippling paralytic disease is about to make an unwelcome return. Poliovirus has re-emerged on Europe's southeastern flank — in Israel and Syria — leaving public-health officials concerned that the disease could be imported and again become established on the continent.

Europe is surprisingly vulnerable. The World Health Organization (WHO) declared its European region, which now spans 53 countries from Portugal to Russia, free of polio in 2002. But many countries have since dropped their guard. Surveillance systems are often incomplete and of poor quality, and suboptimal vaccination rates mean that many countries, including the United Kingdom and Germany,

are considered to be susceptible to outbreaks sparked by imported cases^{1,2} (see 'Polio threat').

The situation is "a wake up call", says Marc Sprenger, director of the European Centre for Disease Prevention and Control (ECDC) in Stockholm. Given the weaknesses in European polio defences, the extensive levels of travel between Europe and Israel, and the millions of refugees fleeing Syria, the ECDC thinks that there is a real risk of outbreaks in the European Union (EU). Member states are taking the threat of import "extremely seriously", Sprenger adds.

The effort to eradicate polio has made great strides since the launch of the Global Polio Eradication Initiative in 1988. Then, 350,000 children were paralysed annually in 125 countries. That toll has been slashed in the past 25 years by more than 99%, with just 223 cases last year. Polio is now endemic in only three countries: Afghanistan, Nigeria and Pakistan.

Sporadic imported cases continue to occur, however, particularly in Africa, with a Somali outbreak causing 174 cases so far this year.

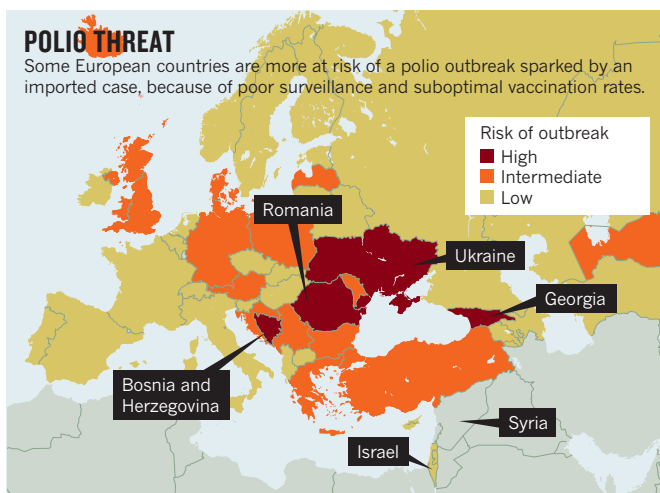
The latest threat emerged on 19 October, when the WHO reported a cluster of cases of acute flaccid paralysis — a classic polio symptom — in Deir-ez-Zor, a conflict-ridden province in eastern Syria. Two of the 22 cases were confirmed as polio by national authorities, and on 29 October, the WHO confirmed a total of ten. Officials had assumed the worst. "Everyone is moving to outbreak-response mode," Oliver Rosenbauer, a spokesman for the Global Polio Eradication Initiative at WHO in Geneva, Switzerland, told *Nature* last week.

The cases are probably a result of a steep fall in child immunization rates in Syria owing to the ongoing war. Because there tend to be about 200 non-paralytic cases of polio for every paralytic one, the cluster is probably "only the ►

► tip of the iceberg”, says Sprenger. There is a major risk that the disease could become endemic in Syria, adds Rosenbauer.

Israel faces a different but also concerning situation. It has high levels of child immunization against polio, but wild poliovirus has been found in sewage in several towns in southern Israel since February. The virus has also been detected in the West Bank and Gaza. The ECDC and the WHO estimate a high risk of international spread of poliovirus from Israel, given the prolonged circulation of virus over a large area.

Israel has so far identified 42 people shedding poliovirus in their faeces. None of them had symptoms of paralysis, and they had been fully vaccinated with inactivated poliovirus vaccine (IPV), which is used in routine immunizations and protects against all polio strains. This is the first time that widespread wild polio has been found without any clinical cases. Most EU countries use IPV, and if exposed to imported polio they could be faced with silent spread of the virus in the environment too, putting unvaccinated populations, particularly infants, at risk. IPV gives a high level of individual protection, but provides poor gut immunity, meaning that vaccinated people might still shed the virus in



faeces. An alternative is oral poliovirus vaccine (OPV), a weakened form of the live virus that provides strong gut immunity and prevents faecal shedding. It is used for mass vaccinations and outbreak control because it is effective, cheap and easy to administer. But in rare cases it can cause polio, so polio-free countries prefer to use IPV, which carries no such risk.

To stop silent transmission, Israel has since August given OPV to more than 890,000 children, and Syria has begun administering OPV to 2.4 million children. The WHO and the United Nations Children's Fund (Unicef) are

planning a vaccination campaign in all neighbouring countries. The appearance of polio “is going to have implications beyond Syria”, says Rosenbauer.

Israel's effective sewage-surveillance systems were able to detect the virus before any clinical cases occurred, but in Europe, only a handful of countries monitor sewage. Surveillance for acute flaccid paralysis is also often poor^{2,3}. The risk that imported cases could go undetected and spread before causing outbreaks is very real, says Sprenger. Europeans who are vaccinated would be protected. But in many countries, including Ukraine, Romania and even some

richer nations, polio vaccination rates can be suboptimal. Up to 12 million EU children are not vaccinated against polio.

“We need to improve environmental surveillance and not wait until we have a clinical case of polio,” says Sprenger. ■

1. ECDC. Suspected outbreak of poliomyelitis in Syria: Risk of importation and spread of poliovirus in the EU (ECDC, 2013).
2. ECDC. Wild-type poliovirus 1 in Israel — what is the risk to the EU/EEA? (ECDC, 2013).
3. WHO. Report of the 27th Meeting of the European Regional Certification Commission for Poliomyelitis Eradication (WHO, 2013).

SOURCE: REF 3

GENETICS

Root of maths genius sought

Entrepreneur's 'Project Einstein' taps 400 top academics for their DNA.

BY ERIKA CHECK HAYDEN

He founded two genetic-sequencing companies and sold them for hundreds of millions of dollars. He helped to sequence the genomes of a Neanderthal man and James Watson, who co-discovered DNA's double helix. Now, entrepreneur Jonathan Rothberg has set his sights on another milestone: finding the genes that underlie mathematical genius.

Rothberg and physicist Max Tegmark, who is based at the Massachusetts Institute of Technology in Cambridge, have enrolled about 400 mathematicians and theoretical physicists from top-ranked US universities in a study dubbed 'Project Einstein'. They plan to sequence the participants' genomes using the Ion Torrent machine that Rothberg developed.

The team will be wading into a field fraught with controversy. Critics have assailed similar projects, such as one at the BGI (formerly

the Beijing Genomics Institute) in Shenzhen, China, that is sequencing the genomes of 1,600 people identified as mathematically precocious children in the 1970s (see *Nature* **497**, 297–299; 2013). The critics say that the sizes of these studies are too small to yield meaningful results for such complex traits. And some are concerned about ethical issues. If the projects find genetic markers for maths ability, these could be used as a basis for the selective abortion of fetuses or in choosing between embryos created through *in vitro* fertilization, says Curtis McMullen. A mathematician at Harvard University in Cambridge, Massachusetts, and a 1998 winner of the prestigious Fields Medal, McMullen was asked to participate in Project Einstein and declined.

Rothberg is pushing ahead. “I'm not at all concerned about the critics,” he says, adding that he does not think such rare genetic traits could be useful in selecting for smarter babies. Influenced by a college class he took from a pioneer in artificial intelligence, and by the

diagnosis of his daughter with tuberous sclerosis complex, a disease that can cause mental retardation and autism, Rothberg has long been interested in cognition. He is also in awe of the abilities of famous scientists. “Einstein said ‘the most incomprehensible thing about the Universe is that it is comprehensible,’” he says. “I'd love to find the genes that make the Universe comprehensible.”

There is precedent to the concept of sequencing extreme outliers in a population in the hunt for influential genes. Scientists have used the technique to sift for genes that influence medical conditions such as high blood pressure and bone loss. Some behavioural geneticists, such as Robert Plomin at King's College London, who is involved with the BGI project, say that there is no reason that this same approach won't work for maths ability. As much as two-thirds of a child's mathematical aptitude seems to be influenced by genes (Y. Kovas *et al. Psychol. Sci.* **24**, 2048–2056; 2013).

But other geneticists say that intelligence is so complex a trait, influenced by so many genes, that not even a sample of thousands of people would yield sufficient statistical power to get at its genetic basis. They point to studies such as one published in May that examined the genomes of more than 125,000 people and found only three genetic markers with a small effect on how long an individual stays at school (C. A. Rietveld *et al. Science* **340**, 1467–1471; 2013). Project Einstein “is unlikely to have any statistical power”, says geneticist Daniel MacArthur at Massachusetts General Hospital in Boston who, with colleagues, has amassed a pedigree of 13 million related people to try to tackle the heritability of complex traits (see *Nature* <http://doi.org/ppj>; 2013).

Some participants in Project Einstein are intrigued by the chance to learn about their own genetic sequences, which will be shared with them. “As a science-fiction fan, I like the idea of having my own genome sequenced,” says David Aldous, a mathematician at the University of California, Berkeley. “Maybe I’ll print a segment onto a T-shirt.”

Others say they wouldn’t be surprised if the study found that maths aptitude was not born so much as made. “I feel that the notion of ‘talent’ may be overrated,” says Michael Hutchings, a mathematician also at Berkeley. He adds that even if genetic markers are found, they could be used for good — not to pre-select for maths geniuses, but to help parents to understand the particular abilities of their children and give them the support they need.

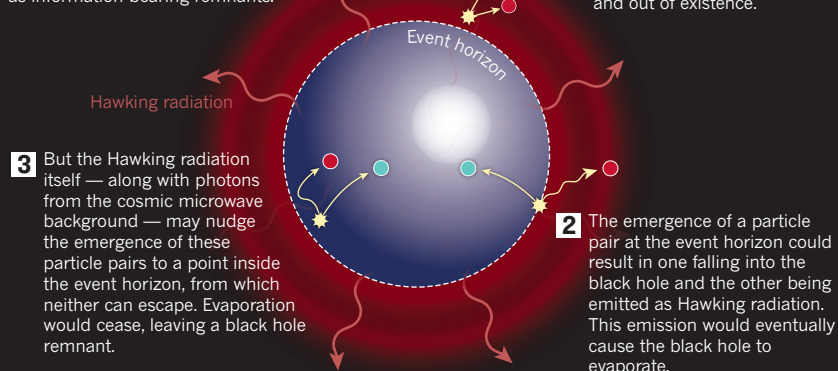
However, McMullen is concerned that the project is appealing to participants’ self-interest without disclosing enough information. “I thought it was strange that it was called ‘Project Einstein’, which seemed designed to appeal to the participants’ egos,” he says. He asked the project’s staff and the New England Institutional Review Board, which approved the study, to explain how results would be used. “The uniform answer to my questions was that ‘we are not responsible for how the information is used after the study is completed,’” he says.

The Rothberg Institute for Childhood Diseases, Rothberg’s private foundation based in Guilford, Connecticut, is the study’s sponsor. But Rothberg won’t say who is funding the project, which other geneticists estimate will cost at least US\$1 million. Some speculate that Rothberg is funding it himself. In 2001, *Fortune* estimated his net worth to be \$168 million, and that was before he sold the sequencing companies he founded — 454 Life Sciences and Ion Torrent, both based in Connecticut — for a combined total of \$880 million.

Rothberg is adamant that the project is well worth the time and the money, whoever is paying for it. “This study may not work at all,” he says — before adding, quickly, that it “is not a crazy thing to do”. For a multimillionaire with time on his hands, that seems to be justification enough. ■

VANISHING ACT

A theorist suggests that black holes may not evaporate, but rather remain as information-bearing remnants.



ASTROPHYSICS

Black holes shrink but endure

Theorist’s idea takes on information–preservation problem.

BY RON COWEN

Old black holes never die, they just fade away. So says veteran cosmologist George Ellis of the University of Cape Town in South Africa, who suggests that the cosmos may be littered with an untold number of shrunken black hole remnants.

Ellis’ speculative report, posted on 17 October on the preprint server arXiv (G. F. R. Ellis <http://arxiv.org/abs/1310.4771>; 2013), seems to undermine the seminal work of Stephen Hawking, a cosmologist at the University of Cambridge, UK. In 1974, Hawking calculated that, owing to quantum effects, black holes are not entirely black: some particles escape the black hole’s gravitational barrier, known as the event horizon. For a solar-mass black hole, these particles, known as Hawking radiation, would be emitted over the course of 10^{67} years until the object vanished without a trace (S. W. Hawking *Nature* **248**, 30–31; 1974).

Although many physicists are sceptical about Ellis’ work, it highlights a long-running debate over the ultimate fate of black holes. Uncertainties abound because of the difficulties in reconciling quantum theory — which predicts the Hawking radiation — and Einstein’s classical theory of gravitation, which defines a black hole’s structure. “This is very much a living issue that people are confused about,” says Vijay Balasubramanian, a string theorist at the University of

Pennsylvania in Philadelphia.

The debate also touches on one of the most cherished beliefs about the Universe: that information is always preserved. If black holes evaporate, then the information they contain may die along with them. By contrast, a black hole remnant would offer a way in which information might be preserved (even if it could never be extracted). By serving as a storehouse, a remnant “could be a fantastic way of resolving all the issues we have with black holes”, says theoretical physicist Jeff Murugan of the University of Cape Town.

In Hawking’s original view, quantum theory permits large fluctuations in energy for brief moments of time. As a consequence, the vacuum of space seethes with particle–anti-particle pairs that continually pop in and out of existence (see ‘Vanishing act’). When this occurs at the event horizon, one member of a particle pair could be sucked into the black hole, whereas the other could escape as Hawking radiation, depleting the black hole’s mass.

But Ellis says that this radiation has another effect. According to Einstein, any source of mass or energy distorts space. A black hole, a body so massive that space closes in on itself, is an extreme example of that distortion. The Hawking radiation would add even more distortion, Ellis says, and so, too, would the ubiquitous photons from the cosmic microwave background, the bath of radiation left over from the Big Bang. He says that these two ►

► types of radiation bend space-time in such a way that the region from which the Hawking radiation is generated eventually gets pushed back behind the event horizon. Once it has been relocated, the radiation can no longer escape and the black hole plateaus at a constant mass.

The analysis is more of an essay than a full-blown study, says string theorist Samir Mathur of Ohio State University in Columbus, because Ellis does not perform a thorough calculation for the bending effect of the radiation. Other physicists say that Ellis is probably incorrect. Nonetheless, Balasubramanian says, the paper is an example of physicists' innate dissatisfaction with evaporating black holes. "The stuff radiates in a weird way, information is lost and then we turn intellectual somersaults to try to account for why the information isn't lost," he says.

But black hole remnants do not offer a perfect solution to the problem of information loss, either. To contain all the information originally stored in a large black hole, the tiny remnants would need to have an infinite number of internal states — which would violate quantum theory, says Mathur.

In 1997, Mathur found a potential way around this problem. He and his colleagues used string theory, a way of marrying quantum theory and gravity, to describe all of the possible states of the gravitational field of black holes. They found that these states lay just outside the classical event horizon in a 'fuzzball' of quantum states. If the fuzzball was the black hole's true boundary, then there would be no Hawking radiation emitted from the event horizon, no evaporation and no information loss, he says.

Last year, others proposed a more violent approach to the problem of information loss. They suggested that the particles in the Hawking radiation did not behave randomly but were instead 'entangled' with each other in such a way that they could be messengers from the darkness, conveying information that had been stored in the black hole. But that entanglement had its own consequence: an enormous amount of energy would be unleashed at the event horizon, such that anything falling in would be instantly immolated (see *Nature* **496**, 20–23; 2013).

The ongoing struggle to preserve the precepts of quantum theory without losing information may offer clues about how and where quantum physics begins to significantly modify Einstein's theory of gravity, says Matt Visser, a cosmologist at Victoria University of Wellington. The conventional picture holds that quantum theory makes big corrections to gravity only well inside the event horizon, near the black hole's singularity — the point at which the density of matter becomes infinite. But some physicists think that quantum physics may be blurring the sharp boundary of the event horizon itself. Ellis's work, Visser says, puts a stronger spotlight on such speculations. ■



KEITH KENT/SPL

Cloud-to-cloud lightning can be detected with sensors, and used as a proxy for intense storms.

METEOROLOGY

Lightning network tested out in Guinea

Project to forecast African storms provides cheap alternative to radar-based weather services.

BY JEFF TOLLEFSON

Meteorologists watched as afternoon thunderstorms brewed in the mountainous region of central Guinea. By the evening of 22 October, the storms had intensified and were moving west towards the coast of Africa. At 8.20 p.m., the meteorologists received a thunderstorm alert, and for the next 45 minutes the 130,000 residents of the city of Fria were hammered by heavy rain, flash floods and winds of up to 77 kilometres per hour.

What happened that evening was not unusual. Similar storms blow through Fria and Guinea's coastal capital Conakry regularly during the rainy season. Flash flooding is a common problem, and the country is frequently buffeted by tornadoes.

What was unusual was the way the storm was detected. Government meteorologists in Guinea lack the Doppler radar system that is usually used for this, and have struggled to track weather using rudimentary equipment. Europe and the United States provide free satellite data and forecasts, but these are

coarse and infrequent. Only in recent months has Guinea turned to a new, simple proxy for storms: flashes of lightning.

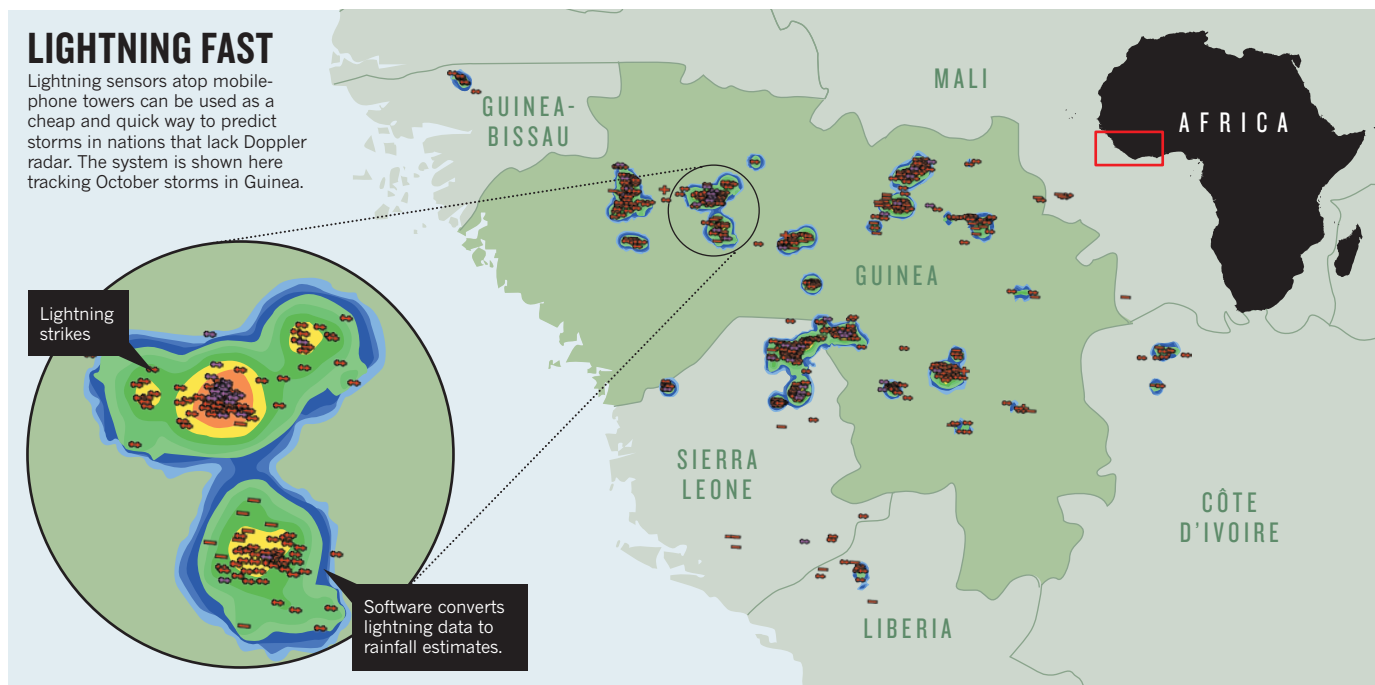
Lightning-detection sensors installed atop just 12 mobile-phone towers now allow the country's meteorological service to track storms nationwide as they develop (see 'Lightning fast'). "With this project, we receive virtually real-time data throughout the entire country," says Mamadou Lamine Bah, director of the service.

The project shows how lightning detection could provide a quick and relatively cheap way for poor countries to acquire basic weather services. Earth Networks, based in Germantown, Maryland, spent around US\$1 million to deploy the current network. By comparison, Bah says, even a simple Doppler radar system would cost around \$10 million. And beyond that initial outlay, a lack of expertise in countries such as Guinea makes those systems prohibitively difficult to maintain and operate.

Earth Networks is currently giving Guinea its lightning data for free, but eventually hopes to sell its services there. Already, the company

LIGHTNING FAST

Lightning sensors atop mobile-phone towers can be used as a cheap and quick way to predict storms in nations that lack Doppler radar. The system is shown here tracking October storms in Guinea.



has contracts worth more than \$5 million with US government bodies, including the National Oceanic and Atmospheric Administration (NOAA) and the Air Force, to provide lightning data from nearly 700 sensors across North America and the Caribbean. It also has smaller customers, such as golf courses and airports. "Our system will work in any part of the world and instantaneously produce results," says Robert Marshall, Earth Networks' president and chief executive. "That's the real breakthrough here."

The technology has its roots in research carried out at the Massachusetts Institute of Technology in Cambridge, and depends on a 26-centimetre-long antenna that registers high-frequency electromagnetic signals from lightning discharges. Earth Networks acquired the technology in 2006, and developed the software necessary to triangulate among multiple sensors to pinpoint individual flashes.

The company has competitors, including Vaisala, based in Vantaa, Finland. But Earth Networks was the first, in 2009, to deploy a system for tracking not just the major lightning bolts that strike the ground but also the more common cloud-to-cloud lightning strikes that are generated as storms intensify.

Earth Networks uses the data to assess storm activity and estimate rainfall, and produces severe-storm alerts 9–16 minutes faster than those generated by NOAA's National Weather Service using Doppler radar.

NOAA is now looking at ways to incorporate lightning data into high-resolution weather models to improve storm tracking. "It's a pretty powerful technology," says Alexander MacDonald, director of NOAA's Earth System Research Laboratory in Boulder, Colorado. Such data will initially come from Earth Networks' tower antennas, but MacDonald says that the agency will eventually collect lightning data from its next geostationary weather satellite, scheduled for launch in 2016.

But the technology's biggest impact could be in countries with limited access to weather services based on satellites and radar. Earth Networks has more than 50 antennas covering most of Brazil and 50 more covering all of India. The company says that it has contracts worth a total of several million dollars in the two countries. The project in Guinea is intended to showcase the technology in a country that

"It's a pretty powerful technology."

has little scientific capacity and hardly any meteorological infrastructure at all.

John Snow, an atmospheric scientist at the University of Oklahoma in Norman and author of a forthcoming World Bank report on weather services in developing countries, says that the Guinea project is an attractive model. The phone towers, which also have basic weather sensors, do not provide ideal conditions for weather stations, Snow says, but the telecommunications company with which Earth Networks has partnered can provide security, technicians and communications. Although custom-built weather stations might provide more-precise data, using phone towers is cheaper and more reliable, he says. "It is a working solution that certainly moves us in the right direction."

Bah says that his agency must now find ways to disseminate the alerts, perhaps by phone. It currently broadcasts radio forecasts just two or three times a day. It must also find money to eventually pay Earth Networks and to formally incorporate lightning data into its operations. But in just a few months, Guinea has moved from rough forecasts to real-time storm tracking. "What I can say from our evaluation is that this system can be deployed in countries like Guinea," says Bah. ■



MORE ONLINE

VIDEO OF THE WEEK



Mouse species' resistance to pain of scorpion venom hints at potential drug target
go.nature.com/zq7gim

MORE NEWS

- PubMed opens post-publication comment platform go.nature.com/dgftwk
- Astronomers will train on simulated data for dark-energy survey go.nature.com/79ekb5
- Atmospheric patterns predict extreme US weather go.nature.com/tjsuha

NATURE PODCAST



Treating chronic HIV; time travel in fiction; and a book on physics under the Nazi regime nature.com/nature/podcast



The eight 40-centimetre telescopes of MEarth are hunting for Earth analogues around M-dwarf stars.

ASTRONOMY

Astronomers revisit dwarf stars' promise

Kepler data spur searches for habitable planets around small, low-temperature stars.

BY EUGENIE SAMUEL REICH

An Earth-like planet — and with it, perhaps life — does not have to orbit a Sun-like star. That idea has inspired astronomers to widen their search for habitable exoplanets to systems surrounding M-dwarfs — stars that are much smaller, redder and cooler than the Sun. Although a project begun in 2008 found just one planet near such a star, data from NASA's Kepler telescope suggest that astronomers should not give up on the approach, and several M-dwarf campaigns are now adjusting their searches.

"I want to know if life can exist on a planet like Earth in a different environment," says astronomer Zachory Berta-Thompson at the Massachusetts Institute of Technology in Cambridge. "That would be one of the most fascinating astrophysical experiments."

For scientists hunting exoplanets that could support life, M-dwarfs offer several advantages. The lower a star's temperature, the closer to the star is its 'habitable zone' — the region in which liquid water can exist on a planet. That means that a habitable planet will orbit an M-dwarf closely, and make more frequent transits across

the star's face — the telltale blockages of light that can be detected from Earth. During transits, some of the starlight is altered by its passage through a planet's atmosphere, and this effect would be proportionally larger for an Earth-sized planet orbiting a star much smaller than the Sun. Astronomers hope that future missions, especially NASA's James Webb Space Telescope (JWST), due to launch in 2018, could detect this light and tease it apart into its spectral components to reveal evidence that gases in the planetary atmosphere are being altered in a way that could signify life.

Motivated by these advantages, astronomer David Charbonneau at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, began a pioneering project to look for 'other Earths' around M-dwarfs close enough to Earth that the light from their atmospheres could be detected with currently planned technology, such as the JWST. The effort, MEarth (pronounced 'mirth'), in 2009 located an uninhabitable M-dwarf planet, GJ 1214b, that was roughly 13 parsecs from Earth and twice its size. The find raised expectations of a huge bounty to come, but four years later, MEarth has not found another M-dwarf

planet, let alone a habitable one. "Personally, I was disappointed," says Berta-Thompson, who worked on the project.

Now researchers think that they understand why the initial search foundered, after examining data collected by NASA's planet-hunting Kepler mission before it ended in August. Kepler surveyed hundreds of thousands of stars, including roughly 4,000 M-dwarfs, and found just 20 planets of the size that MEarth can detect — about twice as big as Earth. But planets the size of Earth or smaller seem to be more common; Kepler found 17 of this size, even though they were much harder for it to see, suggesting that the MEarth team could find habitable planets if it tweaks its search strategy.

In December, Charbonneau will open MEarth's second station of eight telescopes in Chile, to complement a first observatory in Arizona. He plans to concentrate the telescopes on fewer stars — a few hundred rather than a few thousand — for longer time periods. He hopes that spending more time on each target will allow MEarth to detect the subtler signals of the smaller planets.

The project now has competitors. In 2012, a European effort called APACHE (A Pathway Towards the Characterization of Habitable Earths) began taking data from an observatory at Nus in the Italian Alps. Because it can only detect planets about twice the size of Earth or larger, APACHE is likely to see only a handful of planets near its target M-dwarfs. But its principal investigator Alessandro Sozzetti, at the Turin Astrophysical Observatory in Italy, insists that it will be a useful check on the planetary statistics produced by Kepler. "It's still significant," he says.

Another project hoping to take advantage of M-dwarfs is SPECULOOS (Search for Habitable Planets Eclipsing Ultra-cool Stars), a European Research Council mission that will begin taking data on the coolest versions of those stars in 2016. It will use larger telescopes than MEarth (1 metre compared with 0.4 metre) and aims to find planets around smaller stars. Its principal investigator, astronomer Michaël Gillon at the University of Liège in Belgium, says that contrary to some astronomers' assumptions, his preliminary data on some 35 M-dwarfs suggest that their brightness variations are small enough for scientists to detect any Earth-sized planets transiting past them.

Although looking at M-dwarfs has its advocates, it is far from mainstream. Planet-hunting researchers still hope to find a closer analogue to the Sun and Earth. NASA's next big mission is TESS, the Transiting Exoplanet Survey Satellite, a space telescope that will spend two years identifying planets around the 500,000 brightest stars seen from Earth. TESS, which will launch in 2017, will cover many types of stars, including some 10,000 M-dwarfs. That leaves the smaller ground-based campaigns a few more years in the spotlight before TESS takes over. ■

DAN BROCCIOUS/CFR

AGRICULTURE

Farmers dig into soil quality

Analytical technique promises to match fertilizers to soil in bid to boost yields in Africa.

BY QUIRIN SCHIERMEIER

Efforts to bring chemical fertilizers to sub-Saharan Africa are often met with concerns over harmful environmental and economic side effects. The products improve soil quality — useful in Africa, where soils often lack crucial nutrients — and help to increase yields. But fertilizers are costly for subsistence farmers and can leach into water supplies, leading to health problems. Soil scientists have long wanted to assist poor farmers with their selection and use of fertilizers so that they can better match them with soil types and use less of them.

At the Global Soil Week event in Berlin this week, researchers are presenting analytical tools that could enable them to do just that. Rapid soil-fertility assessment, a new spectroscopy technique used to analyse samples and produce site-specific soil maps for farmers, might help to halve the amount of fertilizer needed to bring food self-sufficiency to sub-Saharan Africa, according to proponents. Critics, however, argue that the technique is impractical and costly, and that organic fertilizers such as manure are cheaper and greener.

Malnourishment affects roughly 223 million people in sub-Saharan Africa — about one-quarter of the region's total population — according to estimates by the Food and Agriculture Organization (FAO) of the United Nations. Most of the region missed out on the 'green revolution' that occurred in the late 1940s to the 1960s, which boosted crop yields in other parts of the world. Farmers in some areas of Africa have begun to adopt fertilizers only in the past decade. And, on average, they use just 9 kilograms per hectare each year, compared with more than 200 kilograms used in high-intensity agriculture, such as in Germany. Yields are typically 30–80% lower as a result.

Widespread use of chemical fertilizer has greatly increased food sufficiency in many countries, for example in China and India, but the ecological and health effects of fertilizer chemicals raise serious concerns. China, in particular, is facing a major pollution problem from overuse of nitrogen fertilizer, which can harm air quality and contaminate groundwater, suffocating aquatic life.

To avoid similar problems on African farms, scientists working with local farmers are exploring crop modelling and nutrient management. Field trials to assess the ecological efficiency of organic and chemical fertilizers in different geographic and climatic settings are under way in Ethiopia, Kenya, Malawi, Uganda and Tanzania.



Farmers in sub-Saharan Africa must often contend with soils that lack nutrients for crop production.

The goal, says Rolf Sommer, an agricultural scientist at the International Center for Tropical Agriculture (CIAT) in Nairobi, is to provide smallholders with easy-to-understand information that will help them to decide what works best, economically and ecologically. To help, researchers at CIAT developed the spectroscopy technique, which uses mid-infrared light, to precisely determine the mineral properties, nutrient content and organic chemistry of sub-Saharan soils. Although still in its infancy, the technology could one day be used to evaluate and map the yield potential of soils across the region — information that would then be relayed to farmers. "If farmers know what fertilizer to use and when to use it, they'll need less and it will work better," says Sommer.

Agronomists, he adds, would also be able to make informed recommendations on the right mix and most economical use of fertilizers — a practice that would help to avoid the ecological and monetary costs of fertilizer overuse.

"We do need to increase food productivity in a sustainable way," says Ronald Vargas, a soils and land management officer at the FAO in Rome. "Understanding the health and condition of tropical soils is a prerequisite for any wise use of fertilizers." Advances in soil spectroscopy are "very promising," he adds, and spectral analysis is on the agenda of a special FAO workshop on soil monitoring in Rome this December.

Others are unconvinced. Although soil analysis is "very important," says Johannes Kotschi, a soil scientist at the Association for Agriculture and Ecology in Marburg, Germany, "the notion that intricate and expensive lab technology will

influence farming practices in sub-Saharan Africa is far-fetched". Soil nutrients and soil acidity can be easily determined with readily available, cheap test strips, he adds.

Kotschi also thinks that organic fertilizers such as manure, compost and plant residues should play a greater part in efforts to increase productivity. "Nitrogen fertilizer isn't a quick fix to soil problems," he says. "Its use produces greenhouse gases and it ruins soil fertility rather than improving it." He points out that many smallholders in sub-Saharan Africa can scarcely afford to buy mineral fertilizers anyway.

Most agricultural scientists acknowledge that applying only chemical fertilizers is not the solution to Africa's yield problem. But an organic-only solution, favoured by environmental groups such as Greenpeace, will also fall short, says Sommer.

"Conservation agriculture is excessively time-consuming, and organic fertilizer produced on-farm does not replenish soils with nutrients," he says. "We will need to find solutions that fit local farmers. The times when African farmers blindly applied what Western scientists thought was good for them are definitely over." ■

CORRECTION

The News story 'Final word is near on dark-matter signal' (*Nature* **502**, 421; 2013), wrongly quoted Jonathan Feng as saying the LUX results could rule out all types of neutralino — he believes that only some types would be ruled out.

SHEHZAD NOORANI/STILL PICTURES/ROBERT HARDING

CAN THE SOUTHERN AFRICAN LARGE TELESCOPE LIVE UP TO ITS POTENTIAL?



SOUTHERN STAR

BY LINDA NORDLING

On 30 August, the Cape of Good Hope lived up to its original name, the Cape of Storms. The cargo ship *Atacama* barely managed to find shelter in her berth at Cape Town before the harbour was slammed by towering waves, icy squalls and gale-force winds. A day passed before the tempest subsided and the ship could safely unload its cargo: an automobile-sized device known as the High Resolution Spectrograph (HRS).

It was a fitting welcome. The HRS's final destination, some 350 kilometres inland, was the Southern African Large Telescope (SALT) — a facility that has weathered many storms of its own. Officially completed in 2005, SALT is only now finishing its second year of normal science operations, and pressure is mounting for the facility to prove itself. Along with the

Square Kilometre Array (SKA) of radio telescopes (see *Nature* **480**, 308–309; 2011), SALT is a major component of South Africa's effort to establish its scientific reputation and inspire a new generation of African scientists. Yet its teething problems have prompted questions about its design, the way it was built and how it has been managed so far.

SALT's defenders counter that problems could not have been avoided in building one of the world's largest telescopes on a shoestring. The telescope's first-generation suite of detectors wasn't even complete until the arrival of the HRS. And the spectrograph still needs to be tested and calibrated before it opens for routine use early next year. Nonetheless, Ted Williams, director of the South African Astronomical Observatory (SAAO), which manages SALT, knows that excuses won't suffice. "Our challenge in the coming years is to produce the

nice science that will make people recognize that it's a major telescope that's living up to its potential," he says.

BUDGET BUILD

From the moment planning began in 1998, SALT has been an underdog — not least because the monies available to the SAAO and its partners were never anything like those at the disposal of northern observatories. The SALT team had to find ways to design, build and equip the 11-metre telescope for just US\$30 million — a fraction of the cost of flagship instruments such as the European Southern Observatory's 8.2-metre Very Large Telescope (VLT) in Chile or the 10-metre Keck Telescopes in Hawaii, whose price tags were well in excess of \$100 million. "If the VLT is like a Ferrari, SALT is more like a family car," says Darragh O'Donoghue, head of instrumentation at the SAAO.

JANUS BRINK

Among other things, SALT's location also put it at a technical disadvantage. The site — a desolate hilltop near the town of Sutherland in South Africa's Northern Cape province — was chosen both because it is beautifully dark and because it already had roads and power. The SAAO had been operating a 1.9-metre telescope there since 1974, along with several smaller instruments that had been relocated from their original sites near Johannesburg and Cape Town because of light pollution. But the site stands just 1,798 metres above sea level, quite low by astronomical standards, so there would always be a comparatively thick layer of atmosphere above the telescope to blur the incoming starlight. SALT's images, although acceptable, would never achieve anything close to the clarity of those taken at world-class sites such as the 4,200-metre summit of Mauna Kea in Hawaii.

Nonetheless, SALT's location would allow it to fill a major gap in observations of the southern sky. When a star, galaxy or planet was at or below the horizon for the existing Southern Hemisphere observatories, most of which are in Australia or Chile, it would still be high overhead in South Africa. That is a big plus for astronomers such as Dimitar Sasselov of Harvard University in Cambridge, Massachusetts, who studies exoplanets (those outside the Solar System) by watching for 'transits': tiny, brief dips in a star's brightness as a planet orbits in front of it. "We are often faced with time-critical events like planet transits that make scheduling of observations in the Southern Hemisphere tricky," says Sasselov. So he sees SALT becoming an important complement to large telescopes elsewhere.

Recognizing the limitations they faced, SALT's designers decided to make their telescope world class in the realm of spectroscopy, which involves splitting the incoming light into its constituent wavelengths. Spectroscopy can provide a wealth of information about the composition and motion of celestial objects, but does not require the ultimate in crisp imagery — just a big mirror able to collect lots of light.

The trick was to build such a mirror on a very tight budget. The SAAO accomplished this by copying the radical cost-saving design of the Hobby–Eberly Telescope (HET), an 11-metre-wide instrument completed in 1996 at the McDonald Observatory near Fort Davis, Texas. One big cost saving for this instrument came from a technique pioneered in the 1980s at Hawaii's Keck Observatory. This involved piecing together the telescope's mirror from many hexagonal segments of glass, each of which could be manufactured and shipped to the site independently — a much cheaper proposition than trying to fabricate, polish and transport a single giant mirror.

Another big saving came from designing the dish-like mirror to have a spherical curvature. Spherical mirrors cannot bring starlight to a

sharp focus; that requires a slightly different shape known as a paraboloid. But paraboloids are much more complex and costly to make from segments because each one has a curve that depends on its position. A spherical mirror's segments are identical — meaning that SALT's 91 hexagonal mirror elements could be mass-produced relatively cheaply. And the imperfect focus could be fixed by passing the light through an optical system called a spherical aberration corrector (SAC) before sending it on to the detectors (see 'Inside SALT').

"IF THE VERY LARGE TELESCOPE IS LIKE A FERRARI, THE SOUTHERN AFRICAN LARGE TELESCOPE IS MORE LIKE A FAMILY CAR."

However, as SALT was taking shape, its designers decided to tweak the design of the SAC and other elements to achieve better image quality, which had turned out to be chronically poor at the HET. The changes also allowed for a bigger field of view. But once the construction of SALT was completed in 2005, the telescope faced problems of its own.

The most serious was that, despite the design improvements, SALT's image quality was poor. In images produced by SALTICAM, its main optical camera, and the Robert Stobie Spectrograph (RSS), stars often looked smudged or stretched. However, on rare days, for no obvious reason, the images were perfect.

Solving this mystery took the project team several years. The first potential culprit was SALT's main mirror. The team wondered whether some of the mirror segments were imperfectly aligned, which would prevent them from sending light to the SAC with the required precision. It took nine months of testing the segments to rule out the mirror as the source of the image-quality problems.

The team then moved on to the second suspect: the revamped SAC. This contains auxiliary mirrors up to half a metre in diameter, which needed to be hung to micrometre precision. Checking and testing these mirrors took more than a year.

In the end, the culprit turned out to be the way the SAC was attached to the telescope. The support structure was made out of metals that expanded and contracted differently with changes in temperatures, which affected the alignment of the SAC and explained the variability from day to day. And, during design, the weight of the telescope's instruments had increased, so there had been pressure to keep

the overall weight of the SAC down. This had resulted in the use of a design too flimsy to keep the SAC correctly positioned.

In an ideal world, the SALT team would have foreseen these problems, says Phil Charles, who was the SAAO's director until 2011 and now heads the astronomy group at the University of Southampton, UK. But, given the real-world constraints on time and money, they simply assumed that the Texas group had got this element of the design right. "We now know that HET suffered from all the problems that SALT did," says Charles. "It's just that their basic image quality was so bad that you couldn't see the underlying problems in the mechanical design."

In the end, it was not until August 2010 that a new and better support structure could be hoisted up and mounted on SALT, putting an end to the device's imaging nightmares.

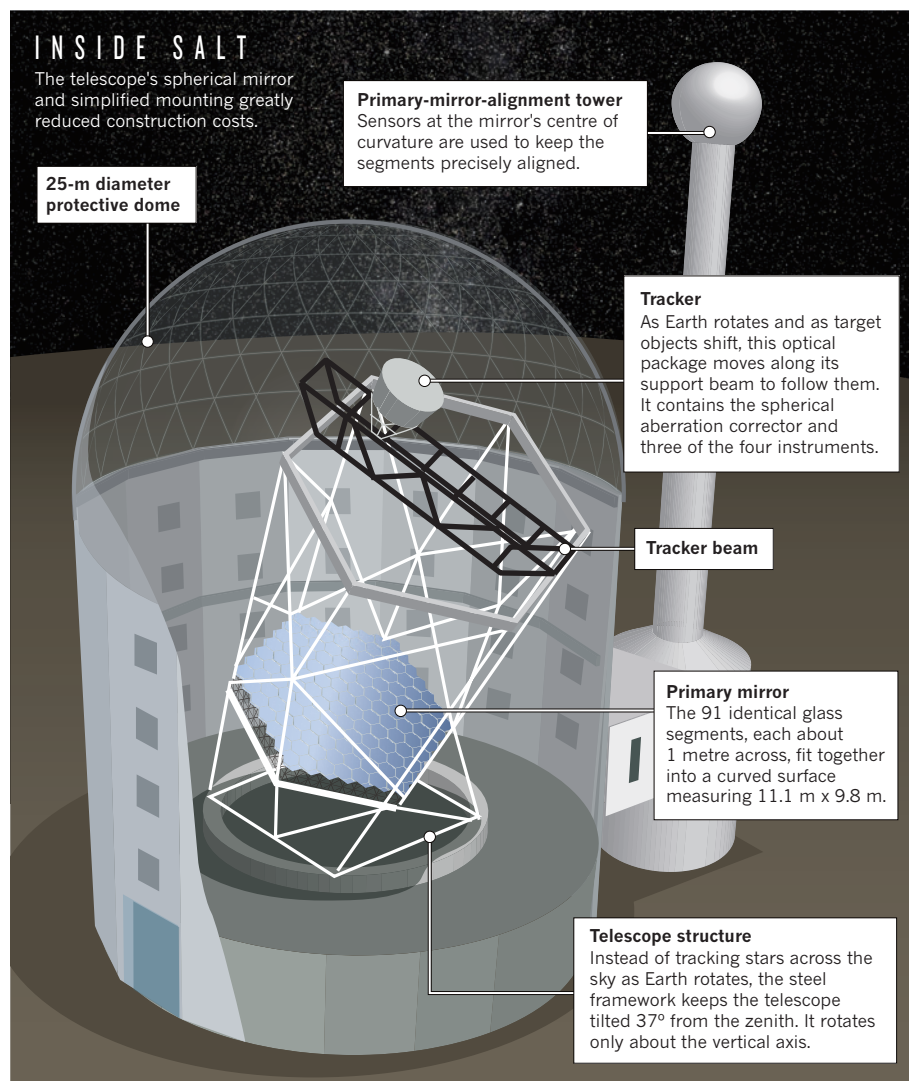
While all this was going on, the team also had to grapple with another major problem, this time with the RSS. Built by one of SALT's major partners, the University of Wisconsin–Madison, it was designed to be the world's most sensitive spectrograph in the near-ultraviolet (UV) range. But, once mounted on the telescope, it managed only average performance at best.

Yet another investigation revealed the culprit to be the spectrograph's plumbing. Because the instrument's large lenses expand and contract with temperature changes, they are mounted in a liquid that is contained in a system of plastic ducts and bladders. The manufacturers of the fluid and the plastics thought they had confirmed that the two would not react chemically, but react they did — producing contaminants that were particularly absorbent in the coveted near-UV range. The spectrograph's optics were dismantled and sent back to the United States, where they were cleaned and the fluid replaced with a different substance. Meanwhile, the cleaning process cracked one of the lenses, so that also had to be replaced. "That set us back as well," says Williams.

COLOUR BLIND

Given this history, it is not surprising that nerves were taut during the recent arrival and assembly of the HRS. And the assembly team soon discovered a problem with the instrument's 'blue camera', which captures and analyses blue light. In effect, the HRS was blind in one eye, recalls Jürgen Schmoll, an optics specialist who had flown in to help with the assembly from Durham University, UK, where the HRS was built.

Closer inspection of the blue camera showed that two of the three metal balls holding its detector chip to micrometre precision had come loose in transit, tilting the chip. The team deliberated over what to do. The safest, slowest and most expensive option was to take the blue camera to a dust-free room in Cape Town and fly a technician in from Durham to perform



the delicate surgery required to reattach the balls. In the end, however, the team decided to fix the problem on-site — a risky operation that involved opening the camera, gluing the balls onto screws, manoeuvring them back into place, breaking the screws loose and cleaning the glue off the balls before closing the camera again. Any speck of dust that entered the camera during this process could become a permanent fixture on the sensitive detector.

A tense few days followed, but on 22 September SALT team members were finally able to declare the operation a success. “The sun’s back out, the spring flowers have thawed ... & we have a healthy blue camera again. Life is good,” they trumpeted on the project’s blog.

Once the HRS is fully up and running next year, the SALT team hopes that it will play an important part in restoring the telescope’s global image. For example, the HRS should be very good at detecting the subtle shifts in a star’s spectrum that show it is being orbited by an exoplanet — and should even be able to study the chemistry of the exoplanet’s atmosphere. Upcoming projects will study the distribution of dark matter in galaxies, and

explore the mechanisms by which galaxies form and grow.

But to stand a chance of playing in the big leagues, SALT needs to pick up its productivity, says Dennis Crabtree, an astronomer at the National Research Council Canada in Victoria who keeps tabs on the scientific output of telescopes larger than 3.5 metres. So far, he says, “SALT has had a pretty low profile because of its teething problems, which led to a low production rate of scientific papers”.

Could the SALT project have avoided those problems? Probably not all of them, says Matt Mountain, the astronomer who oversaw the construction phase of the US Gemini project, which built twin 8-metre telescopes in Hawaii and Chile at an overall cost of \$184 million. Technical glitches are par for the course when building the world’s biggest telescopes, he says. The big question is how these problems are dealt with.

In the past, Mountain explains, when astronomers worked on telescopes roughly 1 metre across, fixing problems as they arose was fairly straightforward. But with today’s giant telescopes, Mountain says, the astronomers need

project leaders and experts from industry to fix problems. It has been a painful transition, he says, from the old astronomer-led system to the new one in which project managers are in charge. But South Africa will need to master that shift to complete its share of the SKA project, which is estimated to cost around US\$2 billion and will involve building hundreds of radio dishes in the African and Australian outbacks.

Williams says that SALT was built through a mixture of the old approach and the new. For both the telescope and the HRS, construction was undertaken with strong project management. However, SALTICAM and the RSS were constructed through the more conventional, astronomer-led approach, he says.

SALT’s supporters remain optimistic, despite the telescope’s technical setbacks. O’Donoghue says that it has fulfilled many of its goals, including placing South Africa on the global astronomy map. “I firmly believe that the SKA would have never come to South Africa if it hadn’t been for SALT as a technology demonstrator,” he says. As evidence of SALT’s influence, he points to the recent decision by the International Astronomical Union to set up an Office for Astronomy Development at the SAAO in Cape Town.

The telescope has also become a technological icon. “Between 2001 and 2003, an average of 50 people visited Sutherland annually,” says Sivuyile Manxoyi, the SAAO’s head of education and outreach. “In 2012 we had 12,205 visitors.” Tourism in the town has also benefited. Today there are 40 guest houses, as well as more than a dozen guest farms, most of which offer stargazing as an attraction. Astronomy is also booming at the country’s universities.

Meanwhile, Williams is trying to raise SALT’s scientific profile by updating Sutherland’s older, smaller telescopes, making it easier for astronomers abroad to carry out observations there by Internet connection (rather than by travelling to South Africa), and by making sure that the science done at SALT is presented at important conferences around the world.

But the output so far from the HET suggests that SALT may never be as productive as the more expensive, multi-purpose telescopes. In 2011, for example, only 26 papers were published on work at the HET, whereas mature telescopes averaged 85, according to Crabtree’s database.

But such gloomy predictions don’t seem to daunt the SALT team. After all, says one of its software engineers, the personality of SALT is “a South African upstart, doing things better and more cheaply than the competition”. Besides, as Williams points out, wintry weather doesn’t usually last long in Africa — and “pessimists don’t build telescopes”. ■

Linda Nordling is a freelance writer in Cape Town, South Africa.

LADY OF THE LAKES

Diane Orihel set her PhD aside to lead a massive protest when Canada tried to shut down its unique Experimental Lakes Area.

BY HANNAH HOAG



It was an ominous way to start the day. When she arrived at work on the morning of 17 May 2012, Diane Orihel ran into distraught colleagues. Staff from Canada's Experimental Lakes Area had just been called to an emergency meeting at the Freshwater Institute in Winnipeg. "It can't be good," said one.

As a PhD student working on her dissertation — not a staff member — Orihel was not allowed in. But an hour later, she heard the news. Michelle Wheatley, regional director of Fisheries and Oceans Canada (DFO), the federal department responsible for the day-to-day operations of the Experimental Lakes Area (ELA), had dropped a bombshell: owing to budget cuts, the field station would close on 31 March 2013. Staff members should begin removing their equipment from labs and lakes. Wheatley instructed them not to speak to the media.

The closure would strike a blow at the heart of freshwater ecology. The ELA — a field site of 58 freshwater lakes in the boreal forest of northwestern Ontario — has since 1968 been the only facility in the world where scientists can manipulate or even intentionally poison an entire lake to monitor the effects. Work there proved that phosphorus from fertilizers sparks algal blooms; quantified the effects of acid rain; showed how mercury accumulates in fish; documented the release of greenhouse gases from hydroelectric reservoirs; and revealed how the synthetic oestrogen in contraceptive pills feminizes male fish. Orihel herself had spent most of a decade doing summer fieldwork in the lakes.

Orihel had no experience as an activist, and was not comfortable in the spotlight. But she was immune to the gagging order because she did not work for the government. She stepped up, becoming public historian, promoter and defender of the site. "I felt a moral obligation, a responsibility, to be the voice for ELA because the ELA scientists couldn't," she says.

At first, Orihel had hoped to get the closure decision reversed within three weeks. She ended up putting her PhD on hold for six months. By the end of that time, she had become one of Canada's most outspoken defenders of science funding and evidence-based policy.

"She's the poster child for the 'scientist-turned-activist,'" says Katie Gibbs, a biologist who co-founded the science-advocacy group Evidence for Democracy in Ottawa.

COURAGE OF HER CONVICTIONS

Orihel was always a high achiever. By the time she started her PhD in 2007, she had published four papers based on her master's research. "Diane impressed me almost on first meeting," says her PhD supervisor, ecologist David Schindler of the University of Alberta in Edmonton. She also had pluck. Schindler hated to see students treat their advisers as gurus, but Orihel was

not afraid to defend her point of view.

Schindler — who in 1991 won the prestigious Stockholm Water Prize for his work on nitrification and acidification of lakes — was an ELA founding scientist and its first director. When he met Orihel, he was 66 years old and shrinking his research group as he readied himself for retirement. He gave Orihel a small grant to investigate an unusual type of algal bloom. She occasionally spoke to residents' associations and advocacy groups about her research, but generally kept to her work.

Everything changed on that Thursday morning in May. After hearing the news, Orihel got in touch with a local politician she had worked with on algae before, and got a quick lesson on media relations from his communications staffer. She wrote a press release and began collecting quotes from scientists. She briefed the opposition party so that it would know what was going on. By mid-afternoon, she had blasted her announcement to a list of journalists that she found online.

On Friday evening, Orihel's neighbours, concerned by her non-stop work, ordered her over for dinner. But she was preoccupied. "Do you know how to set up a website?" she asked her hosts. With their help, the campaign got a name, Save ELA, and an online home. Over the weekend, Orihel organized dozens of volunteers by e-mail and Skype, and persuaded two postdocs to join her in forming the Coalition to Save ELA. They doled out tasks: fill the website with details of past and present projects; develop a petition to deliver to parliament; organize scientists to send an open letter to the fisheries and environment ministers; write more press releases.

By Sunday, Orihel realized that she would need to take leave from her studies. Just three weeks, she told her supervisors. They "weren't keen," she says. "All she had left was to write up her thesis. And she gave that up to save ELA," says Schindler, who regretted the interruption to Orihel's career, but was proud of her gumption. "It was a remarkable show of spine."

EXPANDING CAUSE

The three weeks passed quickly. By June 2012, more than 11,500 Canadians had signed the coalition's petition and Orihel flew to Ottawa to deliver it to politicians. She packed her trip with speeches, meetings and press conferences. She got the open letter — signed by eight established Canadian scientists, including John Smol, a limnologist at Queen's University in Kingston, Ontario — published in a national newspaper. She persuaded scientific societies from the United States, Japan and Australia to send letters of protest to the government. Orihel capped her tour off with a speech to students at Queen's University. Her campaign had grown bigger than just the ELA. "I am you," she began. "I am shy, quiet,

introverted, and get nervous at public speaking." But in the face of "a government's anti-science, anti-environment agenda," she told the audience, "I need your help."

Orihel's campaign tapped into a growing sense of unease. Many Canadian scientists felt that the government's cost-cutting measures had unfairly targeted science and environmental programmes: the Polar Environmental Atmospheric Research Laboratory in the Canadian High Arctic, for example, was slated for partial closure (it later reopened with two-thirds of its previous budget). Federal environmental assessments had been overhauled, reducing the number and length of evaluations. And government scientists were fed up with a communications policy, quietly put in place four years earlier, that restricted their relationship with the press: researchers had to get permission to speak to journalists, and interview requests were often denied or responded to with government-controlled quotes. Greg Rickford, Canada's current science minister, declined to be interviewed for this story, e-mailing only a general statement: "Our government is committed to

science, technology and innovation and taking ideas to the marketplace."

In July 2012, more than 2,000 people gathered for a 'Death of Evidence' rally on Parliament Hill, to protest against the way the government

"All she had left was to write up her thesis. And she gave that up to save ELA."

was undermining science-informed decision-making. "Diane made the ELA one of the best-known examples of funding cuts in Canada," says Gibbs, who organized that rally.

DFO director Dave Gillis insisted at the time that the government had to make the cut — the ELA closure would save Can\$1.5 million (US\$1.5 million) a year, less than 2% of the nearly \$80 million that the DFO needed to trim by 2015 in austerity measures — and said that the department hoped other organizations, such as universities, could take on the ELA's costs. But some claim that ideology drove the decision: the ELA and other research facilities might produce damning data about the environmental impacts of, for example, extracting oil from Alberta's tar sands or the use of industrially valuable chemicals. "It had nothing to do with money. It was inconvenient data," says Smol. In response, the DFO issued a statement to *Nature* saying in part: "Science is the foundation of the department's business and it will continue to build scientific knowledge about our aquatic environment and fisheries resources to support long-term sustainability and conservation objectives."

By last autumn, things were looking bleak: the idea that the ELA might be run by a consortium of universities had come to nothing.

Orihel's dream of getting the original decision reversed was beyond hope: "I knew hell would freeze over before this would ever happen." Exhausted, she stepped down as head of the Coalition to Save ELA, to return to her PhD. But behind the scenes, she knew, ELA scientists had approached the International Institute for Sustainable Development (IISD), a policy think tank based in Winnipeg, to find another way to keep the site going.

LAST-MINUTE REPRIEVE

It came in the nick of time. By mid-March 2013, just weeks before the planned closure, a work crew had started dismantling old cabins at the ELA, and university scientists were contemplating abandoning their experiments. Then, in April, the IISD announced that it had secured a deal with the Ontario provincial government to keep the site open through the summer. Orihel and her coalition deserve substantial credit, says Matt McCandless, the IISD's project manager for the ELA. By keeping the spotlight on the ELA, they "paved the way for the negotiations," he says.

Ontario later promised to provide up to \$2 million a year to run the ELA; in September, the neighbouring province of Manitoba promised \$900,000 over six years. But the battle is not over. On 1 September, the agreement that allowed scientists to pollute the lakes expired. The site is open, but scientists there cannot legally do their game-changing work. "Even today, the ELA is so far from being saved and functional again," says Orihel. "I don't trust this government to do what's in the best interest of science and the environment and all Canadians. This will drag on and on."

She now has her PhD, but Orihel admits that her actions may have limited her career options: the government, or some universities, might be uncomfortable hiring someone so politically vocal. "She took a risk, but her credibility as a scientist will come from her publications," says Smol. Schindler adds: "If you were to ask me to pick the next leader of the ELA project, I'd pick Diane." Some say that she could have been more effective by working with the government, rather than fighting it aggressively. But Orihel does not regret her approach, despite having had to rewrite her personality to run her campaign: "I made myself act unlike myself."

Now Orihel's life is back to normal: she is looking for a postdoc position and finishing off research papers. But she hopes that her actions will inspire others. "Things are so bad in Canada, right now," she says. "Scientists can see the writing on the wall. They're seeing a need to speak out." Orihel says that she has been encouraged to run for public office. "But I have no political aspirations. I just want to be a scientist." ■

Hannah Hoag is a freelance writer based in Toronto, Canada.

➔ **NATURE.COM**

For more on the ELA's reprieve, see: go.nature.com/no42hf

COMMENT



EARTH What will warming mean for hydropower in the Himalayas? **p.617**

CULTURE Time travel in fiction and science as TV's *Doctor Who* turns 50 **p.620**

BEHAVIOUR Two takes on the roots and reach of the human moral instinct **p.622**

OBITUARY David Hunter Hubel, visual-processing pioneer, remembered **p.625**

SERGIO MORAES/REUTERS



The now-full Jardim Gramacho landfill in Rio de Janeiro, Brazil, received more than 10,000 tonnes of waste per day.

Waste production must peak this century

Without drastic action, population growth and urbanization will outpace waste reduction, warn **Daniel Hoornweg, Perinaz Bhada-Tata and Chris Kennedy**.

Solid waste — the stuff we send down our chutes, discard at work and put on the curb every week — is a striking by-product of civilization. The average person in the United States throws away their body weight in rubbish every month. When waste management works well, we give it little thought: out of sight and, usually, quickly out of mind. Discarded materials are collected, some are recycled or composted, and most are landfilled or incinerated. But the global view is troubling.

In the past century, as the world's population has grown and become more urban and affluent, waste production has risen tenfold.

By 2025 it will double again¹. Rubbish is being generated faster than other environmental pollutants, including greenhouse gases. Plastic clogs the world's oceans and rivers, causing flooding in developing-world cities. Solid-waste management is one of the greatest costs to municipal budgets.

The waste problem is acute in emerging cities. Landfills such as Laogang in Shanghai, China; Sudokwon in Seoul; the now-full Jardim Gramacho in Rio de Janeiro, Brazil; and Bordo Poniente in Mexico City vie for the title of the world's largest. Each typically receives more than 10,000 tonnes of waste per day. Rapidly developing cities such as

Shenzhen in China are adding to the world's 2,000-plus inventory of waste incinerators. With the largest able to process more than 5,000 tonnes per day, concerns over ash disposal, air pollution and costs are rising too.

As city dwellers become richer, the amount of waste they produce reaches a limit. Wealthy societies tend to curb their waste. So as living standards around the world rise and urban populations stabilize, global solid-waste generation will peak.

Just when is difficult to predict. But by extending current socio-economic trends to 2100, we project that 'peak waste' will not occur this century. Unless we ►

► reduce population growth and material consumption rates, the planet will have to bear an increasing waste burden.

URBAN PROBLEM

Solid waste is mostly an urban phenomenon. In rural communities there are fewer packaged products, less food waste and less manufacturing. A city resident generates twice as much waste as their rural counterpart of the same affluence. If we account for the fact that urban citizens are usually richer, they generate four times as much.

As urbanization increases, global solid-waste generation is accelerating. In 1900, the world had 220 million urban residents (13% of the population). They produced fewer than 300,000 tonnes of rubbish (such as broken household items, ash, food waste and packaging) per day. By 2000, the 2.9 billion people living in cities (49% of the world's population) were creating more than 3 million tonnes of solid waste per day. By 2025 it will be twice that — enough to fill a line of rubbish trucks 5,000 kilometres long every day.

Together, the member countries of the Organisation for Economic Co-operation and Development (OECD) are the largest waste generators, producing around 1.75 million tonnes per day. This volume is expected to increase until 2050, owing to urban population growth, and then to slowly decline, as advances in material science and technology make products smaller, lighter and more resource efficient.

Some countries generate more waste than others. Japan issues about one-third less rubbish per person than the United States, despite having roughly the same gross domestic product (GDP) per capita. This is because of higher-density living, higher prices for a larger share of imports and

cultural norms. Waste quantities worldwide can also vary seasonally, by up to 30%, as horticultural and food wastes fluctuate. For example, household waste volumes double in the week after Christmas in Canada.

Waste reduction and dematerialization efforts in OECD countries are countered by trends in east Asia, particularly in China.

“Waste will continue to rise in the fast-growing cities of sub-Saharan Africa.”

China's solid-waste generation is expected to increase from 520,550 tonnes per day in 2005 to 1.4 million tonnes per day in 2025. East Asia is now the world's fastest growing region for waste, a distinction that is likely to shift to south Asia (mainly India) in 2025, and then to sub-Saharan Africa around 2050.

As a country becomes richer, the composition of its waste changes. With more money comes more packaging, imports, electronic waste and broken toys and appliances. The wealth of a country can readily be measured, for example, by how many mobile phones it discards. Solid waste can thus be used as a proxy for the environmental impact of urbanization. Most of a material's impact is through production and use. Less than 5% stems from waste management, which includes emissions from collection trucks, landfills and incinerators.

PEAK WASTE

The rate at which solid-waste generation will rise depends on expected urban population and living standards growth and human responses. In 2012, two of us (D.H. and P.B.-T.) authored a World Bank report, *What a Waste*¹, which estimated that global solid-waste generation would rise from more than 3.5 million tonnes per day in 2010 to more

than 6 million tonnes per day in 2025. These values are relatively robust, because urban populations and per capita GDP can be well forecast for several decades.

Extending those projections to 2100 for a range of published population and GDP scenarios shows that global ‘peak waste’ will not happen this century if current trends continue (see ‘When will waste peak?’). Although OECD countries will peak by 2050 and Asia-Pacific countries by 2075, waste will continue to rise in the fast-growing cities of sub-Saharan Africa. The urbanization trajectory of Africa will be the main determinant of the date and intensity of global peak waste².

Using ‘business-as-usual’ projections, we predict that, by 2100, solid-waste generation rates will exceed 11 million tonnes per day — more than three times today's rate. With lower populations, denser, more resource-efficient cities and less consumption (along with higher affluence), the peak could come forward to 2075 and reduce in intensity by more than 25%. This would save around 2.6 million tonnes per day.

CONVERT AND DIVERT

How can today's situation be improved? Much can be done locally to reduce waste. Some countries and cities are leading the way. San Francisco in California has a goal of ‘zero waste’ (100% waste diversion by reduction and recycling) by 2020; already more than 55% of its waste is recycled or reused. The Japanese city of Kawasaki has improved its industrial processes to divert 565,000 tonnes of potential waste per year — more than all the municipal waste the city now handles. The exchange and reuse of materials connects steel, cement, chemical and paper firms into an industrial ecosystem³.

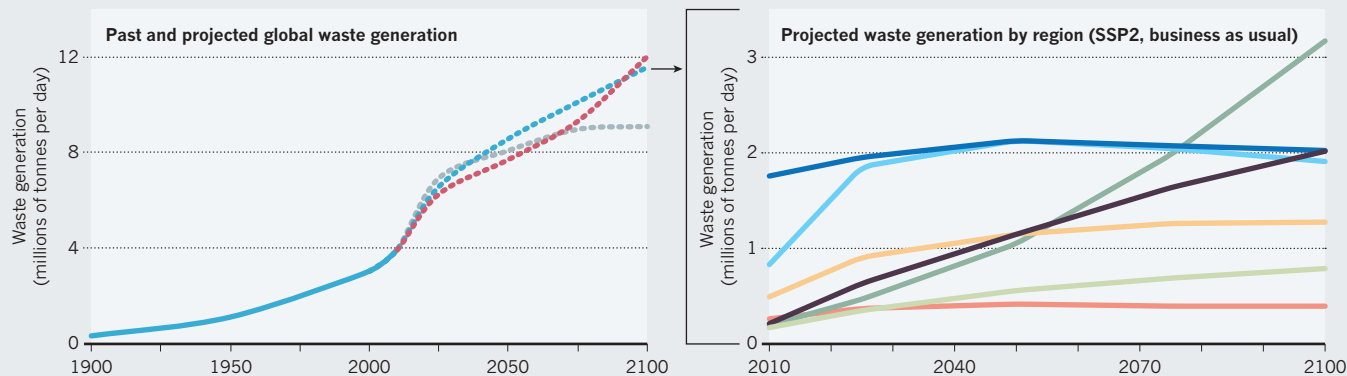
North America and Europe have tried

WHEN WILL WASTE PEAK?

Three projections to 2100 for waste generation spell very different futures. In the first Shared Socioeconomic Pathway⁴ scenario (SSP1), the 7-billion population is 90% urbanized, development goals are achieved, fossil-fuel consumption is reduced and populations are more environmentally conscious. SSP2 is the ‘business-as-usual’ forecast, with an estimated population of 9.5 billion and 80% urbanization. In SSP3, 70% of the world's 13.5 billion live in cities and there are pockets of extreme poverty and moderate wealth, and many countries with rapidly growing populations.

— SSP1 — SSP2 — SSP3

*Organisation for Economic Co-operation and Development



disposal fees, and found that as fees increase, waste generation decreases. Another tactic is to steer people to buy less with their increased wealth, and to spend more on experiential activities that require fewer resources^{4,5}.

But greater attention to consumption and improvement in waste management is needed in rapidly urbanizing regions in developing countries, especially in Africa. Through increased education, equality and targeted economic development, as in the sustainability scenario we evaluated⁶ (SSP1), the global population could stabilize below 8 billion by 2075, and urban populations shortly thereafter. Such a path reflects a move towards a society with greater urban density and less overall material consumption⁷. Also needed is a widespread application of 'industrial ecology' — designing industrial and urban systems to conserve materials. This begins with studies⁸ of the urban metabolism — material and energy flows in cities.

Reducing food and horticultural waste is important — these waste components are expected to remain large. Construction and demolition also contribute a large fraction by mass to the waste stream; therefore, building strategies that maximize the use of existing materials in new construction would yield significant results.

The planet is already straining from the impacts of today's waste, and we are on a path to more than triple quantities. Through a move towards stable or declining populations, denser and better-managed cities consuming fewer resources, and greater equity and use of technology, we can bring peak waste forward and down. The environmental, economic and social benefits would be enormous. ■

Daniel Hoornweg is associate professor of energy systems at the University of Ontario Institute of Technology in Oshawa, Canada.

Perinaz Bhada-Tata is a solid-waste consultant in Dubai, United Arab Emirates.

Chris Kennedy is professor of civil engineering at the University of Toronto, Canada.

e-mail: daniel.hoornweg@uoit.ca

1. Hoornweg, D. & Bhada-Tata, P. *What a Waste: A Global Review of Solid Waste Management* (World Bank, 2012).
2. Dyson, B. & Chang, N. *Waste Mgmt* **25**, 669–679 (2005).
3. Van Berkel, R., Fujita, T., Hashimoto, S. & Fujii, M. *Environ. Sci. Technol.* **43**, 1271–1281 (2009).
4. Ausubel, J. H. & Waggoner, P. E. *Proc. Natl Acad. Sci. USA* **105**, 12774–12779 (2008).
5. Eriksson, O. et al. *J. Cleaner Prod.* **13**, 241–252 (2005).
6. Kriegler, E. *Glob. Environ. Change* **22**, 807–822 (2012).
7. International Solid Waste Association. *Globalization and Waste Management* (2012).
8. Kennedy, C. A. & Hoornweg, D. *J. Indust. Ecol.* **16**, 780–782 (2012).
9. International Institute for Applied Systems Analysis. *SSP Database* (2012); available at go.nature.com/lwp7x1.



The Baltoro Glacier in the Karakoram mountain range feeds the river Indus.

Melting glaciers bring energy uncertainty

Countries should work together to understand how the Himalayan thaw will affect hydroelectric energy, says **Javaid R. Laghari**.

Running 2,000 kilometres from east to west and comprising more than 60,000 square kilometres of ice, the Hindu Kush–Karakoram–Himalayan glaciers are a source of water for the quarter of the global population that lives in south Asia. Glaciers are natural stores and regulators of water supply to rivers, which, in turn, provide water for domestic and industrial consumption, energy generation and irrigation.

Ice cover is decreasing in this region, as for most glaciers in the world, as a result of global warming. Between 2003 and 2009, Himalayan glaciers lost an estimated 174 gigatonnes of water each year¹, and contributed to catastrophic floods of the Indus, Ganges and Brahmaputra rivers. Pollution is accelerating the melt. An 'Asian brown cloud', formed from the 2 million tonnes of soot and dark particles released into the atmosphere every year,

mostly from India and China, warms the air and surface ice².

Seasonal meltwater serves as the main source of power for an increasing number of hydroelectric dams on the rivers served by the glaciers. But hydropower faces a difficult future in south Asia because of climatic, environmental and politico-economic factors. The region is starved of energy, and power shortages of up to 20 hours a day are stunting development. Importing oil and gas from the Gulf, Iran or Tajikistan is expensive or politically difficult. So countries are turning to indigenous hydroelectric power, and to other renewable energies such as solar and wind, for cheap, sustainable energy.

Hydroelectric power must play a part in south Asia's low-carbon energy future. But to be effective, governments around the Himalayas need to work together to measure and model glacier retreat, ►

► changing river flows and their impact on hydroelectric power generation. Political obstacles to dam construction and watershed management must also be overcome.

GLACIAL RETREAT

Glaciers feed thousands of miles of rivers in Pakistan. The largest, the river Indus, depends on glacial waters for half of its flow. But near the river's source, in mountains in the Indian-administered state of Jammu and Kashmir, the glaciers are thinning at an alarming rate of 0.7 metres per year³. The Ganges and Brahmaputra rivers in India and Bangladesh are similarly threatened by glacial melting in the regions of their headwaters⁴.

Modelling of glacier retreat in the Himalayas is hindered by sparse data. Field, satellite and weather records confirm that 9% of the ice area present in the early 1970s had disappeared by the early 2000s (ref. 5). But there has been no comprehensive assessment of current regional mass balance — the difference between the accumulation of ice and its loss³.

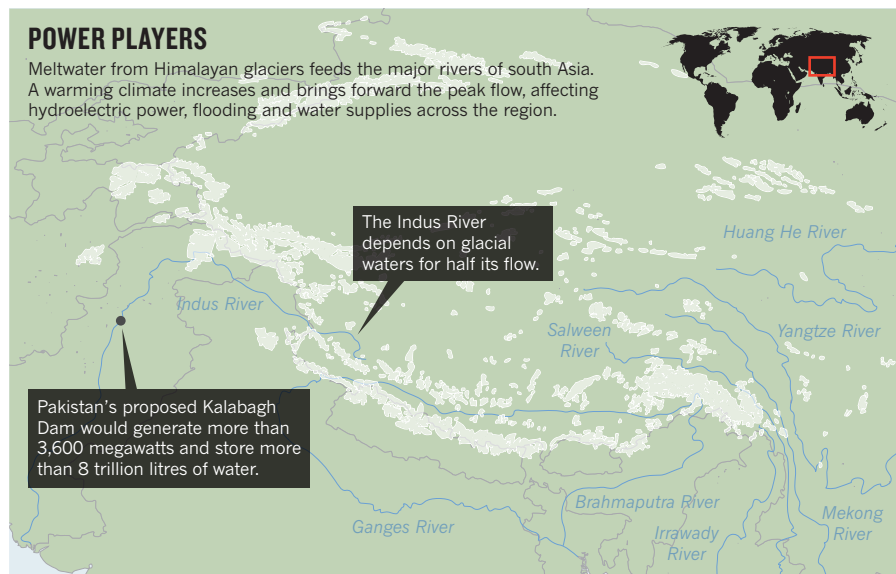
An increased seasonal melt coupled with rains will bring more intense floods, such as those in 2010 caused by excessive monsoon rains that inundated one-fifth of Pakistan's land area for five weeks, killing 2,000 people and costing tens of billions of dollars in damage and economic impact. Sea level is rising at around 3.5 millimetres per year⁵ and the frequency of tropical cyclones is predicted to increase as a result of global warming. Because rain, rather than snow, falls on mountains in spring, river flows will peak before the main growing season. Summers will increasingly see dry streams, withered and abandoned crops, dead fish and low groundwater levels.

With the amount of electricity generated varying with the flow, the changing river patterns in south Asia will disrupt hydropower production. A 1% reduction in stream flow can reduce electricity output by roughly 3% (see go.nature.com/iffhoru). Generation rates will be high in the spring but lower in the summer, when energy is most needed for cooling. Greater uncertainty in the reliability of hydroelectricity will heighten disputes between the countries through which the rivers flow.

Pakistan is currently able to generate 6,700 megawatts (MW) from hydropower on Himalayan rivers, about 37% of the country's total electric-generation capacity. It has the potential to increase this to more than 25,000 MW. India has generation capacity of around 28,000 MW (14%) from hydropower from rivers fed by the Himalayan glaciers and rainwater, and has similar plans for expansion. Nepal and Bhutan currently have 600 MW and 1,500 MW capacity from hydropower, respectively, but have the potential for much more. In 2009, India

POWER PLAYERS

Meltwater from Himalayan glaciers feeds the major rivers of south Asia. A warming climate increases and brings forward the peak flow, affecting hydroelectric power, flooding and water supplies across the region.



and Bhutan signed a protocol in which India will develop 10,000 MW of hydroelectric-generation capacity in Bhutan by 2020 and import surplus power into India.

Hydroelectric power has its challenges: environmental impacts, ecological changes, displacement of population and the politics of the effects of dams on water storage and distribution within and between countries.

For example, the proposed construction of the Kalabagh Dam on the Indus in Pakistan's Khyber Pakhtunkhwa province, which would generate more than 3,600 MW and store more than 8 trillion litres of water, has caused concern in several provinces for more than 30 years (see 'Power players'). Downstream, the Sindh province is concerned that its share of the Indus water will be cut as water stored by the dam upstream goes to irrigate farmlands in Punjab and Khyber Pakhtunkhwa. The coastal mangrove forests of Sindh could become saline deserts without the constant flow of water from the Indus to prevent seawater from encroaching. Khyber Pakhtunkhwa's concern is that large areas of the 1,700 square kilometres of Nowshera District, with a population of more than 1.4 million, would be submerged, and a wider area would suffer from waterlogging and increasing salinity, as has occurred since the building of the Tarbela Dam in the same region.

REGIONAL ACTION

India and Pakistan have been in dispute over India's plans to build 60 dams on the river Chenab in Jammu and Kashmir since the first, the Baglihar Dam, was completed in 2008. Pakistan is concerned that India might block off rivers feeding the Chenab — which flows through Pakistan downstream — especially in a time of war. However, India did not revoke the Indus Water Treaty, signed with Pakistan in 1960, during any of the three subsequent Indo-Pakistani wars.

As the Himalayan glaciers melt, the lack of data for predicting how river flows will change in response is of great concern. South Asian countries urgently need to initiate joint funding and strategies, in collaboration with international, environmental and development agencies, to allow research, data collection, evaluation and monitoring of glacial and climate change to take place.

Himalayan governments and scientists need a strategy for mapping and monitoring the glaciers and deriving mass changes. Academic and scientific communities need to collaborate in research across national borders, and political leaders need to include glacier melting and the future of hydroelectric power generation in their dialogues.

The regional goal should be to maintain stability of flow by taking measures to limit black-carbon emissions, control global warming, slow glacial melting and maintain hydropower generation once its peak is reached, while simultaneously pursuing other clean-energy options, such as wind and solar, for future growth.

Modelling the changing river flows caused by glacial melting and climate change, and their impact on flooding and on power generation must begin immediately so that the future of hydroelectric power in south Asia is safeguarded. ■

Javaid Laghari is chair of Pakistan's Higher Education Commission.
e-mail: jlaghari@hec.gov.pk

1. Gardner, A. S. *et al. Science* **340**, 852–857 (2013).
2. Teng, H., Washington, W. M., Branstator, G., Meehl, G. A. & Lamarque, J.-F. *Geophys. Res. Lett.* **39**, L11703 (2012).
3. Kääb, A., Berthier, E., Nuth, C., Gardelle, J. & Arnaud, Y. *Nature* **488**, 495–498 (2012).
4. United Nations Educational, Scientific and Cultural Organization. *Water in a Changing World* (UNESCO, 2009).
5. Yao, T. *et al. Nature Clim. Change* **2**, 663–667 (2012).



ILLUSTRATION BY DAVID PARKINSON

PHYSICS

The time lord and fellow travellers

As television's time-bending *Doctor Who* turns 50, **Andrew Jaffe** explores time travel in fiction and science.

In 1963, an extraterrestrial burst from his time machine and onto British television screens to the strains of a deliciously eerie electronic theme tune, courtesy of the BBC's pioneering Radiophonic Workshop. *Doctor Who* was born. The series' doughty and eccentric time lord has been zipping from past to future ever since in his TARDIS (Time and Relative Dimension in Space). Regenerated 11 times by new actors — including the latest, Peter Capaldi — the Doctor is going stronger than ever since the show's 2005 reboot.

Generations of fans in numerous countries have fallen for his world-saving, time-hopping antics. But in much science fiction the possibilities of time travel are taken more seriously: its creators try to build a coherent set of rules. Because once you allow the capacity to change past or future, in the real world or in fiction, anything seems possible. For more than a century, time travel has been a rich vein for science-fiction writers and even some scientists — especially those willing to travel farthest from the known laws of physics.

Robert Heinlein's short story "All You Zombies—" is noted for its rigorous internal logic: all the main characters are the same individual at different times in his/her subjective life. The character travels back and forth through time, changing sex and becoming both of his own parents. In a sense, the story is completely self-consistent: cause and effect seem to be preserved, from the protagonist's point of view; life progresses, albeit with science-fictional sex changes and time travel. But something — that is, someone — has been created from nothing, seemingly violating the local laws of physics (and biology).

Amazingly, this kind of time travel is not obviously forbidden by the laws of physics on a global scale. Einstein's general theory of relativity allows 'closed timelike curves' in which a particle can travel back to the same space-time event at which it began. Travelling along such a curve, everything seems fine. But to other observers, something or someone seems just to pop into and out of existence.

This makes physicists very uncomfortable, so theoretical physicist Igor Novikov and collaborators proposed a 'self-consistency principle' in which time travel is possible. Such a trip must be free of paradoxes, and have a single, coherent four-dimensional view of space and time: travel into the past can happen only if it occurred in the Universe's past! Mathematically, we would lose the ability to make predictions in such a Universe, or parts of it: we don't have enough information about the future to know whether a time traveller will emerge in the present. (This also makes physicists very uncomfortable.)

The US television series *Lost* codified the self-consistency

NATURE.COM

For Graham Farmelo on Albert Einstein, see:

go.nature.com/1k2rwy

principle as ‘Whatever Happened, Happened’: even an atomic bomb exploded by the castaways cannot change the past and bring them home. The 1995 film *12 Monkeys* (or its 1962 progenitor *La Jetée*) similarly plays with the chronology of a single event: a character sees himself release an apocalyptic virus and usher in the very future he was sent back in time to prevent.

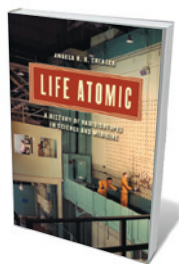
But maybe time travel can occur in other ways; perhaps it is possible to change the past after all. The *Star Trek* episode ‘City on the Edge of Forever’ has Dr McCoy travelling back in time to Depression-era America. There he saves the life of a woman, thus changing the future so that the *Enterprise* is never built. So Kirk and Spock travel back in time and change the past, to save the present, at the price of annihilating the woman with whom Kirk has fallen in love.

In Charles Stross’s 2009 sci-fi novella *Palimpsests* (much influenced by Isaac Asimov’s *The End of Eternity* and 1930s sci-fi pioneer Olaf Stapledon’s grander cosmic visions), each intervention in the past revises present and future. In this scenario, when you kill your ancestor, the Universe becomes one in which you were never born. So you never went back in time, so you didn’t kill your ancestor, so you were born, so you were able to travel through time, so you did kill your ancestor, so ... Stross makes a virtue of this: the initiation into his Stasis, a sort of universal time-police, is to go back and kill your own grandfather.

The other change-the-past trope is to make today’s world a better place by getting rid of some of its more evil past denizens — Desmond Warzel’s short story *Wikihistory* (go.nature.com/txib8y) is the funniest version of this I’ve seen: newbie time travellers keep killing Hitler, so the gurus have to go back and fix the past each time. Or perhaps each intervention cleaves off a new Universe, as in the so-called many-worlds interpretation of quantum mechanics (itself a rich source of scientific and science-fictional ideas, as discussed previously in these pages; go.nature.com/f3oz9w). Indeed, it seems that understanding the possible (or impossible) physics of time travel will require a full understanding of the ‘theory of everything’, marrying general relativity and quantum mechanics.

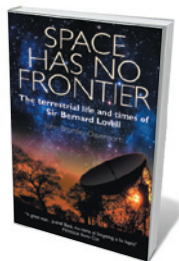
In any scenario, it seems impossible to have time travel without paradoxes or violations of physical laws. So some physicists have theorized that a corollary to the fundamental laws may be that time travel is effectively impossible. In some varieties of the principle, any time machine is censored, hidden inside a black hole formed as a side effect to its creation, walled off from the rest of the Universe by an event horizon. Stephen Hawking has come up with a stronger version, the ‘chronology protection conjecture’: the laws of physics, relativistic and quantum-mechanical,

Books in brief



Life Atomic: A History of Radioisotopes in Science and Medicine
Angela N. H. Creager UNIVERSITY OF CHICAGO PRESS (2013)

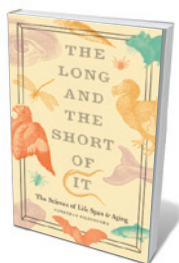
The Manhattan Project’s impact reverberated beyond the atomic bomb, reveals Angela Creager in this lucid scientific history. It paved the way for the Oak Ridge National Laboratory in Tennessee to mass-produce radioisotopes — elemental variants that emit radiation — for peacetime use. These newly abundant products of the “physicists’ war” transformed biology, particularly as molecular tracers in processes such as protein synthesis. Creager deploys radioisotopes as “historical tracers” to explore shifts in medicine, perceptions of cancer risk and the porous “civilian-military divide”.



Space Has No Frontier: The Terrestrial Life and Times of Sir Bernard Lovell

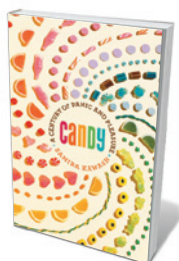
John Bromley-Davenport BENE FACTUM (2013)

He made waves in radio astronomy, founded the UK-based Jodrell Bank Observatory and was an ‘incidental’ cold-war spy. Bernard Lovell, who died aged 98 in 2012, emerges as complex and brilliant in John Bromley-Davenport’s biography. There is much to savour, from Jodrell Bank’s use both in anti-Soviet defence and in tracking the Soviet satellite Sputnik; Lovell’s risky, newly revealed 1963 visit to the Soviet Deep Space Network; and the observatory’s latest role as control centre for the Square Kilometre Array radio telescopes.



The Long and the Short of It: The Science of Life Span and Aging
Jonathan Silvertown UNIVERSITY OF CHICAGO PRESS (2013)

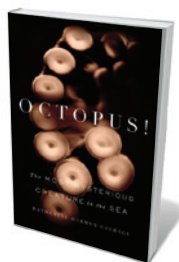
Ecologist Jonathan Silvertown revivifies an old story in this primer on the science of ageing. His look at lifespan centres on a “Methuselah’s menagerie” of bats, naked mole rats, ocean quahogs and humans — in whom cancer is often the price of longevity. He skips from heredity to semelparity (“once-only” reproduction followed by death), drawing on studies of everything from the Japanese hump earwig to human twins. The result is packed with cultural allusions and useful scientific shorthand: if whales lived at the metabolic rate of shrews, for instance, they “would boil the ocean around them”.



Candy: A Century of Panic and Pleasure

Samira Kawash FABER & FABER (2013)

That Halloween haul is a tricky treat. Once reviled as an intoxicant and trigger for lust, candy is now attacked as biochemically dangerous. It is also, as Samira Kawash reveals, a fascinating strand of US cultural history. Sweets evolved from a luxury into the first junk food as, from the 1850s onwards, mass-production technology and sugar chemistry transformed the confectionery industry and built empires such as Mars. Now, argues Kawash, the hidden ‘candification’ of processed foods with corn syrup presents a bigger health hazard than the lollipop — so blatantly sugary that it is easy to avoid.



Octopus! The Most Mysterious Creature in the Sea

Katherine Harmon Courage CURRENT (2013)

Three hearts, eight arms and blue blood — the bizarre appeal of the octopus holds us in a sucker-like grip. They can change colour in three-tenths of a second, thanks to skin sacs called chromatophores. Their arms hold two-thirds of their brain capacity. They can play, use tools, solve mazes and open child-proof bottles. Katherine Harmon Courage’s reportage on what the mollusc is teaching us about robotics, invertebrate intelligence and camouflage is excellent, but sits oddly with the interspersed octopus recipes. [Barbara Kiser](#)

conspire to prevent time machines' construction (or natural occurrence) in the first place.

Some take a more nuanced, if less physically plausible, approach. Stephen King's book *11/22/63* is premised on attempts to change the history of the day on which President John F. Kennedy was shot. "There's a kind of a rule that you'd express as a ratio," King told *Wired* magazine. "The more potential a given event has to change the future, the more difficult that event would be to change." But not all fictional time travel needs to involve material bodies and the problem of causality. In Kurt Vonnegut's *Slaughterhouse-Five*, it is Billy Pilgrim's consciousness that has "come unstuck in time" and travels between upstate New York, the planet Tralfamadore and the firebombed city of Dresden.

More than a century ago, writers were already using time travel for dramatic ends. Mark Twain's *A Connecticut Yankee in King Arthur's Court* skewers the American technophiles of the 1880s as much as the early-medieval Brits whose world he enters. And H. G. Wells's *The Time Machine* is a polemic on the social stratification of late-Victorian Britain, couched in the language of extra dimensions that would inform Einstein's relativistic merger of space and time in the following decades.

In 1899, the playwright Alfred Jarry leapt off from Wells's ideas to make time travel part of his knowingly absurd 'pataphysics', in his pamphlet *Commentary and Instructions for the Practical Construction of the Time Machine*. As part of the Beyond Entropy project with the Architectural Association in London, architect Shin Egashira and I tried to realize some of Jarry's machine. Alas, our success, if any, was aesthetic rather than technological.

And then, of course, there is the time lord himself. Paradoxes rarely trouble the Doctor. Time travel serves mostly as a plot device allowing him to visit humans (much easier on the special-effects budget than aliens) in different circumstances, from the recognizable past to the distant future, defeating his enemies again and again. More recently, however, the show has attempted some sort of cross-temporal continuity, even when this means retroactively changing the past and future to bring his nemeses, the Daleks, back from the dead.

The creators of Doctor Who have tended to favour thrills and chills over scientific (or pseudoscientific) precision. But they have also inspired millions to ponder profound questions about the nature of space and time and our movements through them. Here's to the next 50 years. ■

Andrew Jaffe is a cosmologist at Imperial College London.
e-mail: a.jaffe@imperial.ac.uk

PSYCHOLOGY

The appetite for right

John Whitfield explores two studies that take us from infant ethics to moral choices faced by adults in society.

It would be nice to think that ideas of right and wrong were founded on a blend of insight, experience and instruction. But mostly, instinct is in charge. Morality is an appetite for certain types of behaviour in oneself and others. Like tastes in food and sex, it is rooted in biology, shaped by culture and imperfectly controlled by reason.

As with those other appetites, we develop moral urges because our ancestors prospered by heeding them. Morality underpins social life by guarding against the selfishness that threatens cohesion, and turning that togetherness into a weapon against outsiders. But ethical instincts that put 'us' before 'them' are poorly suited to a globalized world in which different moral systems are in constant contact and problems such as climate change demand cooperation on an unprecedented scale.

Psychologists Paul Bloom and Joshua Greene share this view of the evolutionary roots, social function and limitations of morality. They diverge,

Just Babies: The Origins of Good and Evil

PAUL BLOOM
Crown: 2013.

Moral Tribes: Emotion, Reason, and the Gap Between Us and Them

JOSHUA GREENE
Penguin: 2013.

however, on the aspects of it that they tackle.

In *Just Babies*, Bloom looks at how moral psychology develops in childhood. Using puppet shows or animations that depict helpful or antisocial behaviour, researchers are probing how the ability to judge others develops in infants. These studies, many of which are the work of Bloom and his colleagues, show that ideas of right and wrong begin to emerge so early in life that they must be innate: three-month-olds show that they recognize and prefer good deeds by, for example, looking longer at a kind character than a mean one.

Bloom, ever brisk and authoritative, generally focuses on how things are rather than on how developmental psychology might inform philosophy. His discussion of disgust is particularly good. This is partly because the experiments he describes are nifty. Moral purity, for example, is a value associated with conservative philosophies, and students' political views have been shown to move rightwards when they are standing next to a hand-sanitizer dispenser. And it is partly because he pursues the implications further, arguing that disgust is a poor guide to right and wrong and is liable to make people prejudiced and abusive.

In two senses, Greene picks up where Bloom leaves off. *Moral Tribes* looks at how adults resolve ethical dilemmas, and makes a detailed case for how they should do this.

Greene argues that we have two moral systems that engage different parts of the brain. A fast, automatic, 'tribal' one operates through the emotions and is well suited for solving problems within groups; a slower, deliberative one allows a more impartial perspective. This echoes ideas in Daniel Kahneman's *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011); the relevant bit of Greene's book is even called 'Morality Fast and Slow'. Greene's research has focused on conflicts between the



IMAGE SOURCE/GETTY

two systems, using a philosophical problem called the trolley dilemma, which asks whether it is right to put one person in the path of a runaway tram to save the lives of several more. Most people believe it is wrong to stop the trolley by pushing a man onto the tracks. But most think it is right to flick a switch that diverts the car from a track on which it hits five people onto one where it hits only one.

This seems like an instinctive manifestation of the doctrine of double effect — the philosophical principle that holds that using people as a means to an end is worse than harming them as collateral damage. Greene's experiments, however, seem to show that we think that flicking a switch is more ethical than pushing a person not because of any moral distinction, but because it is socially useful to recoil from physical violence that might provoke retaliation or ostracism.

In place of moral absolutes, Greene carries a flag for utilitarianism. This pragmatic philosophy, developed in the eighteenth and nineteenth centuries by Jeremy Bentham and John Stuart Mill, argues that, to quote Bentham, "it is the greatest happiness of the greatest number that is the measure of right and wrong". The brain's slow moral system, Greene says, naturally arrives at utilitarian decisions, and the philosophy's universality and impartiality transcend faster 'tribal' thinking.

As a science writer who touches on issues beyond science, Greene should be cherished for pursuing his questions wherever they take him and for having the interdisciplinary skills to do so. *Moral Tribes* is clever and absorbing. But although Greene makes a persuasive case for utilitarianism as a means for individuals to live a good life, there is a politics-shaped hole in his suggestion that it might offer a cure for social divides.

It is difficult, for example, to imagine pro-lifers being swayed by Greene's utilitarian argument for the legality of abortion — as the satirist Jonathan Swift remarked, you cannot reason a man out of something that he was not reasoned into. It seems that those who succeed in harnessing a group's tribal instincts tend to defeat those who aim to rise above them. Policy-makers could benefit from reading Greene's book, but anyone with an election to win might be better off with Bloom's focus on morality's automatic weaponry. ■

John Whitfield is the author of *People Will Talk: The Surprising Science of Reputation*.
j.a.whitfield@gmail.com

SPACE SCIENCE

Zero-gravity hero

John Gilbey is gripped by the memoir of Chris Hadfield, a former International Space Station commander.

The third-brightest object in Earth's night sky is the International Space Station (ISS), according to NASA. The station's cultural impact on humanity has perhaps been less brilliant — until this year. From March to May, the tenure of Chris Hadfield as commander of ISS Expedition 35 sparked a worldwide surge of interest in daily life in space.

Hadfield, the first Canadian to walk in space, charmed hundreds of thousands of followers as he tweeted stunning images of Earth rolling beneath him and the gripping and sometimes bizarre minutiae of his day-to-day schedule. (Take this tweet from 8 May: "Yesterday was so cool: as we tested our Soyuz thrusters we could hear and feel them firing, and how they shook and flexed the whole Station.") He even managed an inspired zero-gravity rendering of the Bowie classic *Space Oddity*, complete with guitar. Hadfield brought us a new connectivity with, and understanding of, the work of the ISS crews.

An Astronaut's Guide to Life on Earth describes cogently the core skills that twenty-first-century astronauts need to master — from the unsavoury task of mending the zero-gravity toilet to the challenge of running complex science experiments in orbit. It is clear from the detailed descriptions that working in space remains an enormously complex and routinely dangerous career — Hadfield knew well all seven members of the lost *Columbia* shuttle crew.

Equally compelling is his analysis of the key behaviours required of the aspiring astronaut. The right person fits in with the human and technical environment with the least disruption; a true team member can embed their own skills and expertise in the single entity that is the ISS crew. Only in this way, Hadfield urges, can the apparently trivial and minor everyday faults of such a massively complex system



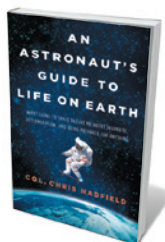
Chris Hadfield prepares for a mission in 2012.

be prevented from escalating into major, life-threatening incidents.

Hadfield's description of his time on the ISS — and the long, complex pathway that took him there — is detailed, frequently technical, amusingly pragmatic and often self-deprecating. The narrative is far from linear — the highlights and weird events pour out in a torrent, leaving you wishing desperately that you had travelled with him.

An Astronaut's Guide to Life on Earth is an impressive memoir of Hadfield's part in developing a permanent home for humanity in Earth's orbit. As the title suggests, the book has many important lessons for those of us destined to remain Earthbound — and especially for those seeking to build a new openness for science and technology through public engagement. Every secondary school student should be given a copy: in terms of inspiration, motivation and a sense of belief in the future of humanity in space, this book ranks alongside the accounts published by the *Apollo 11* astronauts. I can think of no higher praise. ■

John Gilbey teaches in the Department of Computer Science, Aberystwyth University, UK.
e-mail: gilbey@bcs.org.uk



An Astronaut's Guide to Life on Earth: What Going to Space Taught Me About Ingenuity, Determination, and Being Prepared for Anything
CHRIS HADFIELD
Little, Brown: 2013.

Correspondence

Anthropocene: keep the guard up

Chris Thomas writes that recent gains in species numbers associated with climate warming could more than balance species losses (*Nature* **502**, 7; 2013). But conservation is not just about total species richness — it is also about functioning ecosystems.

Unlike ecosystems that have resulted from millennia of competition and predation, we are much less sure about the stability of new, unfamiliar ones created by invasive species, which we know can wreak havoc.

Thomas also extols the virtues of hybrid vigour generated by crosses between native and immigrant species. But again, we have little idea of these newcomers' long-term future because they have yet to be winnowed by selection — in contrast to their contemporary, non-hybrid counterparts. Furthermore, species hybridization flies in the face of conservation, which aims to preserve extant species and their genomes.

We are still ill-equipped to predict the biological effects of climate change. It would therefore be foolish from the standpoint of both ecology and evolution to stop protecting pre-Anthropocene ecosystems and species from the onslaught of climate-driven newcomers.

Tim Caro *University of California, Davis, USA.*
tmcara@ucdavis.edu

Anthropocene: action makes sense

Chris Thomas argues that natural hybridization of invasive and native species could have unforeseen ecological benefits (*Nature* **502**, 7; 2013). However, it is dangerous to underestimate non-native species' potential for ecological and economic damage, which can manifest after decades of remaining

innocuous, or in subtle but enormously harmful ways (D. Simberloff *et al.* *Trends Ecol. Evol.* **28**, 58–66; 2013).

Thomas cites *Spartina anglica* as a potential contributor to future biodiversity. But this plant is so widespread and destructive that it is on the International Union for the Conservation of Nature's list of '100 of the World's Worst Invasive Alien Species'.

Any increase in local biodiversity that might arise from introduced species will be countered by an overall global decrease. Hawaii, for example, once had 114 native bird species: 55 of these are now extinct and 53 have been introduced, so the local change in species number is minimal. Globally, however, the change is a disaster. The introduced species are common elsewhere, but 48% of the islands' native species are gone forever.

The draft European Union legislation for controlling introduced species does not assume that their impact is always negative, as Thomas implies. Species will be risk-assessed, and states will then be obliged to control at most 50 of those confirmed to be harmful.

We maintain that combating invading species after gauging the risk they pose is well justified on scientific grounds and is not "irrational".

Daniel Simberloff *University of Tennessee, Knoxville, USA.*
tebo@utk.edu

Piero Genovesi *Institute for Environmental Protection and Research, and IUCN SSC Invasive Species Specialist Group, Rome, Italy.*

Climate panel is ripe for examination

Sociologists of science wish to study the Intergovernmental Panel on Climate Change (IPCC) for the same reason that they want to examine other loci at which scientific knowledge is made — whether in a laboratory, the field, a museum or at a

conference. We too approached the IPCC in autumn 2010 with a request to study it from the inside; we too were told 'no' (see *Nature* **502**, 281; 2013).

We therefore had to rely on self-reported accounts. Using document analysis and interviews with lead authors, we analysed how authors navigate the distinction between scientific description and value judgements, for example when offering information pertaining to the definition of 'dangerous climate change'.

The IPCC has become a dominant institution in climate science — in the assessment of knowledge for policy-making, and in how assessment practices alter empirical and computer-simulated climate science. Global knowledge assessments such as those undertaken by the IPCC call for carefully documented systematic studies by trained ethnographers.

Let us hope that the IPCC will recognize itself as a legitimate object for scholarly investigation this time around.

Mike Hulme, Martin Mahony *King's College London, UK.*
mike.hulme@kcl.ac.uk

Urban greening needs better data

Current urban-greening programmes are all too often based on inadequate data (see, for example, C. T. Driscoll *et al.* *BioScience* **62**, 354–366; 2012), and models for estimating the value of urban vegetation are largely untested. To make substantive progress towards urban sustainability, city managers and researchers need to know where, when, how and which greening programmes are appropriate for urban areas.

Simplified urban-forest models have been widely used to estimate the benefits of scattered planting of trees in city parks and avenues, but these mostly fail to build in estimates of uncertainty or to consider trade-offs and costs. For

example, urban forests would be unlikely to reduce atmospheric concentrations of polluting particulates and nitrogen dioxide (H. Setälä *et al.* *Environ. Pollut.* **183**, 104–112; 2013), and their high pollen density could exacerbate respiratory conditions such as asthma.

We suggest, therefore, that urban-greening strategies should be tailored specifically to their localities. Programmes need to be validated by testing against comparative studies that capture spatial and temporal variability in and among cities. This means that local urban data collection and ecosystem modelling will have to meet the same high standards as those applied to non-urban areas.

Diane E. Pataki *University of Utah, Salt Lake City, USA.*
diane.pataki@utah.edu

*On behalf of 9 co-signatories (see go.nature.com/blzh2i for full list).

Bird vision offers sharp insight

I find the explosion of interest in the visual system of mice surprising, given that murine eyesight is equivalent to 20/2,000 vision in humans. With their 20/50 vision, pigeons might offer a less "blurry picture" of human 20/20 visual acuity (*Nature* **502**, 156–158; 2013).

Mammals and birds have different brain set-ups, but they both have two, similar visual systems (T. Shimizu and A. N. Bowers *Behav. Brain Res.* **98**, 183–191; 1999). Moreover, pigeons could sidestep the logistical and ethical concerns associated with monkey models of visual processing.

Then there are the untouchables — the birds of prey, which could teach us a thing or two about crystal-clear viewing. The eagle's vision, for example, is widely believed to be about 20/4 in human terms.

Damian Scarf *University of Otago, Dunedin, New Zealand.*
damian@psy.otago.ac.nz

David Hunter Hubel

(1926–2013)

Neuroscientist who helped to reveal how the brain processes visual information.

When David Hunter Hubel died on 22 September, the world lost a great neuroscientist. It also lost a passionate advocate for a style of small-scale research that may still be one of the most powerful ways to make discoveries.

Hubel studied the brain circuitry that underlies vision in collaboration with neurophysiologist Torsten Wiesel. By working on one neuron at a time, the pair gave neuroscientists a new understanding of a cortical circuit that contains millions of neurons and hundreds of millions of connections.

Hubel was born in Windsor, Canada, and grew up in Montreal. His father was a chemical engineer, and, as a boy, Hubel enjoyed tinkering with both electronics and chemistry. In 1947, he graduated from McGill University in Montreal with a bachelor's degree in mathematics and physics. He then took a leap and attended medical school, also at McGill.

Hubel received his doctor of medicine in 1951 and trained as a neurology fellow at Johns Hopkins University School of Medicine in Baltimore, Maryland. His studies were briefly interrupted when he was drafted into the US Army. In 1958, neurobiologist Stephen Kuffler invited Hubel to join his lab and work with Wiesel at the Wilmer Eye Institute at Johns Hopkins. Thus began an extraordinary 25-year collaboration. In fact, Hubel and Wiesel worked so closely and published so extensively together that some people thought that they were one person, named Hubel N. Wiesel.

In 1959, Hubel and Wiesel moved with Kuffler to Harvard Medical School in Boston, Massachusetts. When they started their experiments in the late 1950s, the visual cortex was essentially uncharted territory. In their 2004 book, *Brain and Visual Perception* (Oxford University Press), Wiesel described how he and David “approached the visual cortex as explorers of a new world”.

Kuffler had shown during the 1950s that shining spots of light on a small part of a rabbit's field of view triggered strong signalling in specific neurons in the retina, with nearby neurons having overlapping ‘receptive fields’. This work revealed that, in travelling from the photoreceptors to the output neurons of the retina, visual information is deconstructed into an image that resembles a George Seurat-like pointillistic painting.

But how it is resynthesized to generate a complete picture of the world, as seen by both eyes, remained a mystery.

Hubel and Wiesel found that the spots of light that so effectively activated retinal

that visual experience could alter it. They observed that, during a restricted developmental window, when vision in one eye was impeded, the good eye ‘hijacked’ cortical circuits that should have been shared

equally between the two eyes. This finding was the first example of how experience can change brain circuits. In 1981, Hubel and Wiesel were awarded the Nobel Prize in Physiology or Medicine for their studies of how the visual cortex processes information, along with neurobiologist Roger Sperry.

David had a lifelong love of art and music. Sometimes in the evenings, when I was a PhD student in his and Wiesel's lab during the 1970s, the haunting sounds of his flute would waft down the hallway. He also enjoyed fabricating the tools of his trade, such as tungsten microelectrodes, often using the lathe that he kept in his lab.

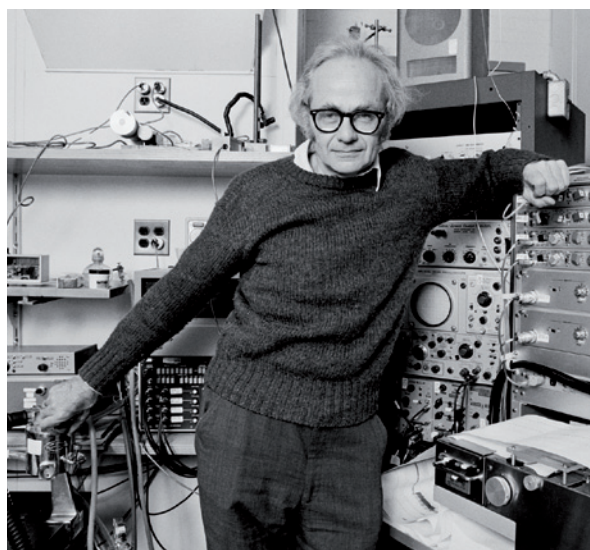
Aside from his children (David had three sons with his wife, Ruth), his largest personal legacy may be his passion for fundamental, discovery-based

research. He communicated this enthusiasm to legions of graduate and medical students with spellbinding lectures, often using illustrations of optical illusions to link the science of the visual system to the beauty of art and visual perception.

David worried publicly about the state of biomedical research: large labs lead by faculty who are too busy with grant-writing and administration to be able to participate in their own experiments. Even after he closed his own lab, he ran a seminar for Harvard undergraduates, teaching them the fundamentals of neuroscience as well as hands-on lab techniques, including how to use a lathe, solder a circuit board and look through a microscope.

David showed generations of young scientists and clinicians how science can become art and how art can become science. Our understanding of the brain and perception has changed profoundly because of him. ■

Carla J. Shatz is director of Bio-X and professor of biology and neurobiology at Stanford University in California. In 1971–1976 she was a PhD student and junior fellow with David Hubel at Harvard Medical School in Boston, Massachusetts. e-mail: cshatz@stanford.edu



neurons and the neurons of the lateral geniculate nucleus (a relay point between the retina and the visual cortex), had no effect on the neurons in the visual cortex. Then, one evening, the pair noticed that a view of the edge of one of their stimulus slides made cortical neurons respond robustly. Furthermore, many of these ‘edge-detecting’ neurons were responding to information from both eyes.

Hubel and Wiesel had just discovered the first visual-circuit steps used to reassemble our binocular view of the world. They reported their findings in two beautifully written studies in *The Journal of Physiology* in 1959 and 1962.

In the years that followed, Hubel and Wiesel found that neurons responsive to the same line or edge orientation clustered together in vertical columns stretching from the outer surface of the visual cortex to its inner white matter. And neurons responding best to stimuli presented to the right or to the left eye were located near each other, also in vertical clusters.

This columnar architecture seemed so precise that initially, researchers thought it was hard-wired. But through studies of cats and monkeys that had had one eyelid sealed, as well as of children with congenital cataracts, Hubel and Wiesel found

Drivers of decoupling in drylands

A study reveals that increasing aridity alters the balance of carbon, nitrogen and phosphorus in dryland soils, providing insight into how global climate change will affect soil fertility and ecosystem services. [SEE LETTER P.672](#)

DAVID A. WARDLE

In all terrestrial ecosystems, the cycling of chemical elements is driven by both biotic and abiotic processes, but the ways in which these factors affect cycling vary for different elements. This is in part because carbon and nitrogen inputs to ecosystems are driven largely by biological processes such as photosynthesis and nitrogen fixation, whereas phosphorus inputs are driven mostly by weathering of rocks. The different driving processes can result in these cycles showing decoupled responses to environmental influences¹, as has been most convincingly demonstrated for long-term chronosequences (gradients of soils of different ages) in which the availability of phosphorus relative to nitrogen diminishes sharply over time^{2,3}. On page 672 of this issue, Delgado-Baquerizo *et al.*⁴ provide compelling evidence for a similar type of decoupled response to increasing aridity for dryland ecosystems worldwide.

The authors analysed soil samples collected from 224 plots in dryland ecosystems that vary greatly in aridity, from all continents except Antarctica. They found that, as aridity increases, there is a decline in concentration of both the total and the most biologically available forms of carbon and nitrogen, but an increase in the most biologically available forms of phosphorus. The decline in carbon and nitrogen most probably emerges through impairment by moisture limitation of biological processes that drive their ecosystem inputs and fluxes, whereas the increase in available phosphorus results from greater weathering of phosphorus-containing rocks and reduced uptake by plants (Fig. 1).

Delgado-Baquerizo and colleagues' results reveal that, because the response of soil phosphorus to increasing aridity is decoupled from that of carbon and nitrogen, there are sharp increases in the ratios of both carbon to phosphorus and nitrogen to phosphorus. Using a statistical approach for estimating causative

The balance between carbon, nitrogen and phosphorus will become increasingly disrupted as ecosystems become drier.

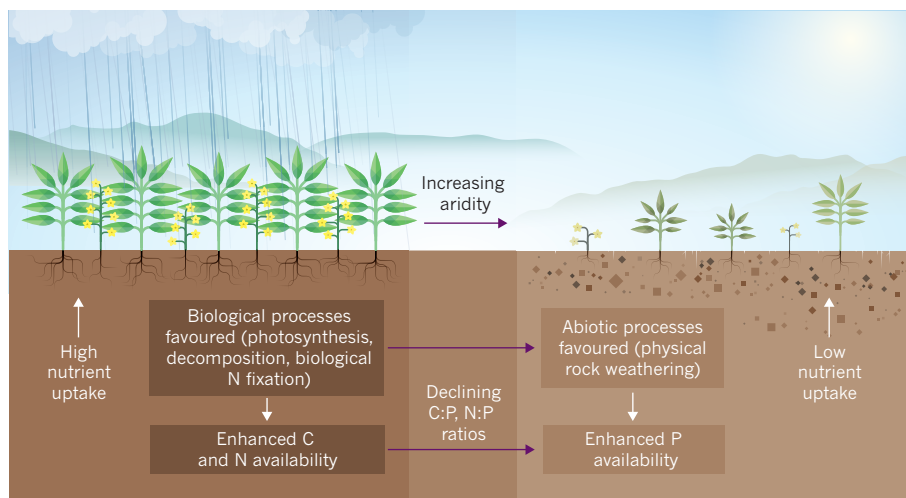


Figure 1 | Aridity causes elemental cycles to decouple. Aridity is predicted to increase in many dryland ecosystems worldwide because of global climate change. Delgado-Baquerizo *et al.*⁴ report that, as aridity increases, available soil carbon (C) and nitrogen (N) decline, whereas available soil phosphorus (P) increases. This is a consequence of the impairment of biological processes that contribute to the C and N levels, and of an increase in the relative importance of abiotic processes that contribute to P availability.

pathways, known as structural-equation modelling, they also provide evidence that aridity has a direct negative effect on soil organic-matter content and phosphatase enzyme activity (an indicator of biological phosphorus demand), but a positive effect on total and inorganic phosphorus. These lines of evidence suggest that, as aridity increases, the growth of plants and other biota should become more limited by nitrogen (because it is cycled in organic matter), but less limited by phosphorus.

There are distinct parallels between the findings of this study and those of previous work that examined ecosystem development on newly formed land surfaces (that is, primary succession). It has been shown for several long-term chronosequences that such new surfaces have a high availability of phosphorus relative to nitrogen, and that as the significance of biological processes increases over time, organic matter in soil develops so that phosphorus becomes increasingly limiting relative to nitrogen^{3,5,6}. The changes that occur with increasing aridity therefore operate largely in the opposite direction to those that occur during ecosystem development. It would thus be interesting to ascertain whether shifts in the

balance of nitrogen to phosphorus resulting from aridity have knock-on effects on vegetation characteristics, ecosystem processes and above-ground and below-ground biota, as has been shown in previous studies of ecosystem development⁵.

The finding that aridity leads to increasing availability of phosphorus relative to nitrogen may at first sight seem paradoxical, especially given that some of the world's driest regions, such as much of Australia, are also among the most phosphorus-limited⁷. Severe phosphorus limitation is especially apparent for very old soils (including those in arid regions) from which phosphorus has been lost over long periods and cannot be replenished. In drylands, as in all ecosystems, there are likely to be multiple drivers of the availability of phosphorus relative to nitrogen, with soil age reducing it^{3,8} and, as shown by Delgado-Baquerizo *et al.*, aridity enhancing it. Although there is some evidence that rainfall patterns influence how nutrient balances respond to soil age⁹, the extent to which the balance between soil nitrogen and phosphorus is driven by soil age rather than by aridity, or by the interactive effect of these two factors, remains a largely open question.

Delgado-Baquerizo and co-workers'

findings offer fresh insight into the consequences of the increased aridity that is projected to occur because of human-driven global climate change¹⁰. Specifically, they reveal that, for dryland ecosystems worldwide, the balance between carbon, nitrogen and phosphorus will become increasingly disrupted as ecosystems become drier. This will occur through widespread losses of soil organic matter (and therefore of biologically available pools of carbon and nitrogen), and through an increased role of abiotic factors (and therefore of available phosphorus). Of particular concern are the authors' data showing that this decoupling of elemental cycles accelerates in a nonlinear manner as aridity increases, suggesting that, as global climate change progresses, the ecosystem properties of many drylands could pass a tipping point that will be difficult or impossible to reverse.

The study highlights the fact that, as aridity increases, adverse ecological consequences will arise not only through the direct effects of moisture limitation, but also through the indirect effects of decoupled elemental cycling and reduced organic matter in soil. This could have far-reaching consequences. For example, dryland ecosystems will be able to store less carbon both above and below ground, thus compromising their ability to mitigate increased levels of atmospheric carbon dioxide and climate change. More immediately, reduced soil carbon and nitrogen may impair the supply of nutrients from the soil and therefore the productivity of crops and livestock, with potentially dire consequences for many of the more than 2 billion people who inhabit dryland regions. This study underscores the fact that increased aridity associated with global change, and its effects on soil nutrient balances, could greatly affect the capacity of drylands to deliver key ecosystem services upon which human well-being depends. ■

David A. Wardle is in the Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden.
e-mail: david.wardle@slu.se

- McGill, W. B. & Cole, C. V. *Geoderma* **26**, 267–286 (1981).
- Walker, T. W. & Syers, J. K. *Geoderma* **15**, 1–19 (1976).
- Peltzer, D. A. *et al. Ecol. Monogr.* **80**, 509–529 (2010).
- Delgado-Baquerizo, M. *et al. Nature* **502**, 672–676 (2013).
- Vitousek, P. M. *Nutrient Cycling and Limitation: Hawaii as a Model System* (Princeton Univ. Press, 2002).
- Wardle, D. A., Walker, L. R. & Bardgett, R. D. *Science* **305**, 509–513 (2004).
- McKenzie, N., Jacquier, D., Isbell, R. & Brown, K. *Australian Soils and Landscapes* (CSIRO Publishing, 2004).
- Selmants, P. C. & Hart, S. C. *Glob. Biogeochem. Cycles* **22**, GB1021 (2008).
- Porder, S. & Chadwick, O. E. *Ecology* **90**, 623–636 (2009).
- Dai, A. *Nature Clim. Change* **3**, 52–58 (2013).

STRUCTURAL BIOLOGY

Pivotal findings for a transcription machine

Crystal structures of the complete RNA polymerase I complex are now revealed. The structures link the opening and closing of this enzyme's DNA-binding cleft to the control of transcription. SEE ARTICLES P.644 & P.650

JOOST ZOMERDIJK

RNA polymerases are intricate molecular machines that transcribe DNA into RNA, combining RNA synthesis with the precise movement of a DNA template across their active site. Eukaryotic cells (those of animals, plants and fungi) have several RNA polymerases, each dedicated to the production of specific RNAs. RNA polymerase I (Pol I) synthesizes the ribosomal RNA component of the cell's protein-producing factories and so is crucial for cell survival, growth and proliferation; malfunction of Pol I can cause cell death or support the unrestrained proliferation characteristic of cancer cells¹. In two groundbreaking papers in this issue, Fernández-Tornero *et al.*² (page 644) and Engel *et al.*³ (page 650) present the first crystal structures of the complete 14-subunit yeast Pol I at

3.0 and 2.8 ångströms resolution, respectively. The structures provide unprecedented insight into Pol I-specific features, potential mechanisms in transcription, and evolutionary conservation of the structures and functions of Pol enzymes⁴.

Structural analyses of bacterial Pol, eukaryotic Pol II and archaeal Pol have detailed the architecture of these enzymes, the interactions between their subunits and their inner workings during transcription⁴. Pol I and Pol III share overall architecture with Pol II⁵. However, each Pol contains specific subcomplexes and features that influence its ability to transcribe a particular subset of genes: whereas Pol I produces rRNAs, Pol II generates messenger RNAs and Pol III synthesizes small non-coding RNAs, including transfer RNAs and 5S rRNA⁶.

Pol I is the most productive of the eukaryotic polymerases. To achieve high-throughput transcription, multiple Pol I complexes transcribe the ribosomal DNA, and these are densely packed along each template and are highly processive (likely to traverse the entire template). The Pol I crystal structures reveal various distinguishing features that have the potential to influence the enzyme's output, partly by facilitating its productive association with the DNA template.

The structures confirm that the zinc-ribbon domain at the carboxy terminus of Pol I subunit A12.2 inserts into, and forms an integral part of, the enzyme's active-site region⁷ (Fig. 1). (By contrast, TFIIS, the functional counterpart of A12.2 in Pol II, only transiently associates with the active site of paused Pol II.) Within the active site, this zinc ribbon can stimulate the removal of faulty and redundant RNA sequences to prevent Pol I arrest and consequent 'traffic jams' along the template, thus increasing transcription efficiency. The stability of A12.2 in Pol I is influenced by its interaction with the (TFIIF-like) dimerization domain of the Pol I-specific subcomplex A49–A34.5 (ref. 8). The structural data now rationalize this, revealing the contact points between A12.2 and the A49–A34.5 amino-terminal dimerization domain, as well as extensive interactions of the A34.5 C terminus as it wraps around the outer face of the A135

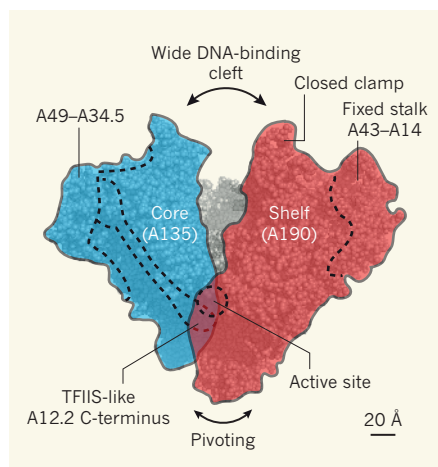


Figure 1 | Structure of RNA polymerase I. Front view of the 14-subunit, 590-kilodalton polymerase I complex with approximate locations of the core and shelf modules, which broadly overlap the cleft sides formed by subunits A135 and A190, respectively, as well as the DNA-binding cleft. The open and closed states of the cleft are determined by relative pivoting of the core and shelf modules at the cleft base near the active site, where RNA synthesis occurs. Also indicated are the locations of: the permanently closed clamp; the fixed stalk, which consists of the A43–A14 subcomplex and contributes to permanent closure of the clamp; the TFIIS-like C-terminal domain of A12.2, which is integral to the active site; and the A49–A34.5 subcomplex, which stabilizes A12.2.

*This article and the papers under discussion^{2,3} were published online on 23 October 2013.

subunit, which help to anchor the subcomplex.

Procession of RNA polymerases along a DNA template is facilitated by a closed-clamp component. The structures of Pol I reveal that the A43–A14 subcomplex, which comprises a fixed stalk (Fig. 1), contributes to a permanently closed state of the clamp and, therefore, to the high processivity of Pol I. By contrast, the clamps of other RNA polymerases are mobile elements. In Pol II, for example, attachment of the Rpb4–Rpb7 stalk locks the clamp in a closed state over the complex of RNA and DNA template during transcription, but this stalk is detachable^{9,10}.

Intriguingly, both teams' crystals are dimers of Pol I, in which the stalk of each Pol I inserts into the DNA-binding cleft of the other Pol I, through the A43 C-terminal 'connector' domain, thus making extensive contacts with the cleft and the coiled-coil motif of the clamp. The dimers have an unusually wide cleft (Fig. 1), perhaps partly because of this A43-connector insertion.

The cleft is too wide to anchor the complex of RNA and DNA template, particularly near the active site. This widening contributes to further rearrangements near the active site. (For example, crucial 'aspartate-loop' interactions are configured differently from those in Pol II; the 'bridge' helix contributing to DNA movement through the active site is unfolded in the middle and kinked; and there is partial blockage of the gate to the exit channel for newly synthesized RNA.) Furthermore, the wide cleft is occupied by a Pol I-specific extended loop of A190, which the authors refer to as the expander³ or DNA-mimicking loop². Because of its location, this loop would interfere with DNA loading at the active site. In one of the three Pol I structures presented by Fernández-Tornero *et al.*, no loop is detectable at the active site, hinting that it is unlikely to be essential for stabilization of the expanded cleft, although not excluding a role in its establishment.

Fernández-Tornero and colleagues' crystals display varying degrees of cleft widening. Comparative structural modelling of RNA polymerases suggests that the Pol I cleft widens as a result of relative pivoting of 'core' and 'shelf' modules (which are formed mainly by the largest subunits, A135 and A190) at the base of the cleft, near the active site^{2,3} (Fig. 1). Engel *et al.* draw parallels to similar domain pivoting in inhibitor-bound or paused bacterial Pol, in which a pivoted or ratcheted state is associated with cleft opening and coupled rearrangements of domains near the active centre, inactivating the polymerase^{11,12}.

Because the new structures imply that the DNA template must be loaded into Pol I that has a closed clamp, perhaps the open or shut status of the cleft contributes to DNA-loading efficiency. It is possible that binding of the DNA template in the open cleft of a Pol I monomer triggers cleft closure, potentially coupled with

relocation of the expander loop, rendering the enzyme active. Cleft closure by pivoting of the core and shelf modules presumably occurs concomitantly with refolding of the bridge helix, opening of the RNA-exit gate and the approach of A135 to anchor the DNA template in the active site. An understanding of the exact rearrangements will hinge on structural analysis of Pol I engaged in transcript elongation and, therefore, in complex with DNA and RNA.

Engel *et al.* propose that regulatory factors binding at the core–shelf interface might facilitate cleft closure. They speculate that Rrn3 (a factor that tethers Pol I to proteins bound specifically to promoter DNA sequences) triggers cleft closure by binding Pol I near the RNA-exit channel^{3,13}. This attractive possibility awaits confirmation, perhaps through analysis of a Pol I–Rrn3 co-crystal. Conversely, factors that terminate transcription by Pol I might induce cleft opening. In all probability, the regulation of transcription by modulation of the core–shelf interface, which is seen in bacterial Pol, is also a feature of eukaryotic RNA polymerases³.

Solving the crystal structure of the complete Pol I complex is a triumph, providing a wealth of information with which to build a picture of the specific mechanisms and control of rRNA-gene transcription in eukaryotes and also to explore the general mechanisms of transcription by all RNA polymerases. Another tour de force will be necessary to solve the structure of Pol I in transcription-elongation mode and, further, that of the

complete Pol I pre-initiation complex, incorporating Rrn3, the core promoter-binding factors (Rrn6, Rrn7 and Rrn11 with TBP) and the rDNA promoter sequences. Such structures, together with those presented by Fernández-Tornero *et al.* and Engel *et al.*, will yield information that is vital for establishing when and where crucial protein and DNA contacts are made, disrupted and rearranged as Pol I steps through the transcription cycle. ■

Joost Zomerdijs is at the Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK.

e-mail: j.zomerdijs@dundee.ac.uk

1. Russell, J. & Zomerdijs, J. C. B. M. *Trends Biochem. Sci.* **30**, 87–96 (2005).
2. Fernández-Tornero, C. *et al.* *Nature* **502**, 644–649 (2013).
3. Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D. & Cramer, P. *Nature* **502**, 650–655 (2013).
4. Werner, F. & Grohmann, D. *Nature Rev. Microbiol.* **9**, 85–98 (2011).
5. Cramer, P. *et al.* *Annu. Rev. Biophys.* **37**, 337–352 (2008).
6. Vannini, A. & Cramer, P. *Mol. Cell* **45**, 439–446 (2012).
7. Kuhn, C.-D. *et al.* *Cell* **131**, 1260–1272 (2007).
8. Geiger, S. R. *et al.* *Mol. Cell* **39**, 583–594 (2010).
9. Armache, K.-J., Kettenberger, H. & Cramer, P. *Proc. Natl Acad. Sci. USA* **100**, 6964–6968 (2003).
10. Bushnell, D. A. & Kornberg, R. D. *Proc. Natl Acad. Sci. USA* **100**, 6969–6973 (2003).
11. Tagami, S. *et al.* *Nature* **468**, 978–982 (2010).
12. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. *Cell* **152**, 431–441 (2013).
13. Blattner, C. *et al.* *Genes Dev.* **25**, 2093–2105 (2011).

QUANTUM PHYSICS

Single electrons pop out of the Fermi sea

The ability to control individual electrons in an electronic conductor would pave the way for novel quantum technologies. Single electrons emerging from a sea of their fellows in a nanoscale electrode can now be generated. SEE LETTER P.659

CHRISTIAN FLINDT

Splashing water in the bath usually leads to small waves, splashes and droplets. Similarly, applying a voltage pulse to the sea of electrons in a nanoscale electrode produces a complex quantum state involving several electrons that have been kicked out of the sea, as well as holes — or missing electrons — left behind. On page 659 of this issue, Dubois *et al.*¹ report the first experimental voltage-pulse generation of just a single electron, not several, emerging on top of an electronic sea*.

*This article and the paper under discussion¹ were published online on 23 October 2013.

A nanoscale electrode is a reservoir of electrons, often referred to as a Fermi sea. Applying a voltage to the electrode amounts to changing the sea level by either pouring more electrons into the electrode (thus increasing the sea level) or emptying out electrons (thereby decreasing the sea level). A voltage that varies with time typically stirs up the Fermi sea and causes waves and splashes of electrons. This effect led Levitov and colleagues^{2–4} to investigate theoretically how a time-dependent voltage affects a Fermi sea. Surprisingly, and quite remarkably, they found that a particular shape of voltage pulses should excite just a single electron onto the surface of the Fermi

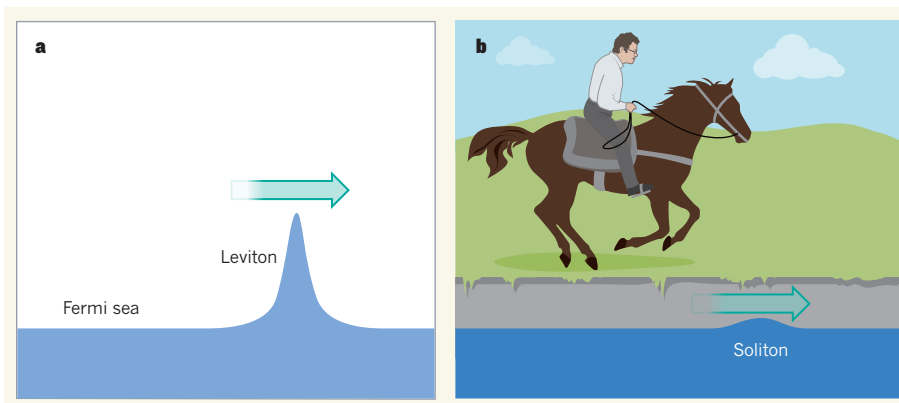


Figure 1 | Spotting solitons and levitons. **a**, Dubois *et al.*¹ have experimentally realized Levitov and colleagues' proposal^{2–4} that a carefully engineered voltage pulse would bring just a single electron to the top of the Fermi sea of electrons in a nanoelectrode; they have named the resulting single-electron wavepacket a leviton. **b**, Levitons resemble solitons, waves that keep their shape while travelling at constant speed and which were first observed in Edinburgh's Union Canal by John Scott Russell while on horseback.

sea, leaving no traces behind. This job would be done by a voltage pulse that changes in time according to the mathematical function called a Lorentzian.

In their experiment, Dubois *et al.* realize the proposal by Levitov and colleagues, and they name the resulting single-electron wavepacket a leviton, because it resembles a soliton in certain ways. Solitons were first observed in the nineteenth century by the Scottish engineer John Scott Russell who noticed that a boat brought to a sudden stop in the Union Canal running into Edinburgh generated a single, localized wave of water that travelled several kilometres without changing its shape or slowing down (Fig. 1). Such self-sustained waves are now known as solitons, and they occur in a variety of systems described by non-linear wave equations.

Just like solitons, levitons of different heights, widths and creation times can be superimposed in a controllable manner and travel unhindered on top of a Fermi sea. To produce levitons, Dubois *et al.* used a nanoscale circuit consisting of two electrodes connected by a small conductor. They applied Lorentzian-shaped voltage pulses on one electrode to generate levitons that travel through the conductor to the other electrode.

Whereas Russell observed solitons in the Union Canal from the back of his horse, the observation of levitons requires sophisticated experimental techniques. To observe them, the temperature must be as low as it can get to make the Fermi sea as quiet as possible. Dubois and colleagues managed to cool their sample down to 35 millikelvin, close to absolute zero. A sequence of Lorentzian-shaped pulses should yield a noiseless flow of levitons without electrical fluctuations^{2–4}. The authors measured the electrical noise⁵ and found only the background noise caused by tiny thermal fluctuations. Next, they used a narrow constriction in the conductor — a quantum

point contact — to filter out a fraction of the levitons. By measuring the increased noise due to the filtering, they could infer the number of emitted levitons and demonstrate that each pulse produces exactly one leviton, with no additional disturbances.

To corroborate their findings, the research team performed a Hong-Ou-Mandel experiment known from optics⁶. Here, a semi-transparent mirror randomly reflects or transmits photons into two different output arms. However, if two identical photons simultaneously hit each side of the mirror, they always exit into the same output arm. The photons are said to 'bunch', as is typical for the class of particles called bosons. Levitons, by contrast, are fermions, which 'anti-bunch' by exiting into different output arms⁷. Dubois *et al.* generated levitons in both electrodes and caused them to interfere at the quantum point contact, which acts as a semi-transparent mirror. Levitons arriving simultaneously at the quantum point contact were found to anti-bunch, confirming their fermionic nature.

Dubois and colleagues' work demonstrates unprecedented control of single electrons in the Fermi sea of a nanoelectrode, and it opens up a plethora of applications and directions for fundamental research. One can envisage future quantum electronics with levitons — levitonics — in which single levitons are emitted into a circuit architecture with edge states (formed in a strong magnetic field) that function as rails for the levitons by guiding them to beam splitters and interferometers for further processing, borrowing ideas and concepts from quantum optics.

Additional experiments might investigate the statistical properties of levitons, including the fluctuations in the number of levitons (full counting statistics⁸) and the distribution of waiting times between levitons⁹. A leviton can contain more than one electron and may even carry just a fraction of the electron charge if implemented in a one-dimensional system



50 Years Ago

During the past year, reports of a remarkable case of 'digital vision' have percolated into Britain from the U.S.S.R. ... The subject ... whose personality is admittedly abnormal, is said to have trained herself to distinguish colours and forms by means of her fingers and to be able to read books and newspapers by digital scanning alone ... It has been shown, for example, that her reading is not impaired by interposing a plate of glass between the print and her fingers or by projecting the print on to a ground glass screen to exclude tactile sensation. It might, therefore, seem that the girl's fingers are genuinely sensitive to light ... The lack of any image-forming device and the relative poverty of the nerve supply to the fingers in comparison with that of the eye constitute seemingly fatal objections to the hypothesis of 'digital vision'.

From *Nature* 2 November 1963

100 Years Ago

Vorlesungen über die Theorie der Wärmestrahlung. By Dr. Max Planck — The first edition of this book, which appeared in 1906, was reviewed in *Nature* ... The many and varied contributions to our knowledge of radiation phenomena that have been published in the ensuing years have made it necessary for Dr. Planck to rewrite and modify the book to a considerable extent ... As before, the object of the book is to apply the statistical methods previously used in the kinetic theory of gases to the phenomena of radiation ... The treatment is largely based on the remarkable assumption which the author designates as the "quantum-hypothesis." ... This is analogous to the electron theory, which assigns a definite magnitude to the electron or "elementary quantum" of electricity.

From *Nature* 30 November 1913

of interacting electrons known as a Luttinger liquid. Atomic levitons may also be realized in Fermi gases of cold atoms. Further down the road, one can imagine solid-state quantum computers with levitons acting as the fundamental carriers of quantum information. The realization of on-demand levitons is a major step forward in the attempts to realize quantum electronics with timed emissions of single electrons into a nanoscale quantum circuit. There are plenty of promising prospects ahead. ■

Christian Flindt is in the Department of Theoretical Physics, University of Geneva, 1211 Geneva, Switzerland.

e-mail: christian.flindt@unige.ch

1. Dubois, J. *et al.* *Nature* **502**, 659–663 (2013).
2. Levitov, L. S., Lee, H. & Lesovik, G. *J. Math. Phys.* **37**, 4845–4866 (1996).
3. Ivanov, D. A., Lee, H. W. & Levitov, L. S. *Phys. Rev. B* **56**, 6839–6850 (1997).
4. Keeling, J., Klich, I. & Levitov, L. S. *Phys. Rev. Lett.* **97**, 116403 (2006).
5. Blanter, Y. M. & Büttiker, M. *Phys. Rep.* **336**, 1–166 (2000).
6. Hong, C. K., Ou, Z. Y. & Mandel, L. *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
7. Bocquillon, E. *et al.* *Science* **339**, 1054–1057 (2013).
8. Nazarov, Y. V. (ed.) *Quantum Noise in Mesoscopic Physics*, NATO Science Series (Kluwer, 2003).
9. Albert, M., Haack, G., Flindt, C. & Büttiker, M. *Phys. Rev. Lett.* **108**, 186806 (2012).

energetic and entropic properties of a system at equilibrium, and the magnitude of the frictional forces that determine how fast the system can move around on its free-energy surface. Any molecule will spend almost all of its time in free-energy valleys (minima of the free-energy surface). In the case of protein folding, these valleys correspond to the folded and the unfolded states (Fig. 1).

The probability of a molecule passing from one valley to another — that is, how frequently the reaction takes place — is dominated by the height of the barrier between the valleys. The most interesting event, however, which contains essentially all of the information about the sequence of molecular steps in the reaction, is the actual crossing of the barrier. The molecules spend only a tiny fraction of their time in this transition-state region, and information about their passage is correspondingly hard to get hold of.

Chung, Eaton and colleagues last year succeeded⁵ in measuring these microsecond transition-path times by recording the fluorescence from individual protein molecules and analysing the signal photon by photon, using a clever, previously reported method⁶. They have now taken these investigations a crucial step further by probing the dynamics of a small helical protein in the transition-state area in unprecedented detail.

For protein folding, information about the structural properties of molecules at the top of the barrier has previously been inferred from investigations of how individual changes in a protein's amino-acid sequence affect its folding kinetics⁷. The timescales of barrier crossing have been studied in laser-induced temperature-jump experiments⁸

BIOPHYSICS

Rough passage across a barrier

The dynamics of chemical reactions in solution are described by Kramers' theory, but the parameters involved have eluded direct measurement. A study of protein folding reveals how this problem can be overcome. SEE LETTER P.685

BENJAMIN SCHULER & JANE CLARKE

Proteins fluctuate between different conformations to perform their sophisticated tasks. The random motion of water molecules around proteins provides an inexhaustible reservoir of thermal 'kicks', which act as the molecular driving forces of such conformational dynamics. Counter-intuitively, the same thermal motions of the solvent also limit the speed of biomolecular motion, an effect known as solvent friction. But it has become increasingly clear that, in some important cases, friction within a protein molecule might be the dominant impediment to its molecular dynamics. In this issue, Chung and Eaton¹ (page 685) report one of the most impressive studies so far in which the contribution of such internal friction to dynamics is quantified for protein molecules caught in the act of folding. Remarkably, the results have implications far beyond protein folding*.

The rate at which chemical reactions proceed is most commonly conceptualized in terms of a barrier-crossing process. In the simplest case considered in most chemistry textbooks, this barrier might correspond to the energy required to break a single chemical bond in a molecule in the gas phase. The generalized concept of barrier crossing, developed by the Dutch physicist Hans Kramers and published² in 1940,

can also be applied to much more complex processes³, including reactions in solution, and even protein folding⁴.

The formulation of such a simplified description of reaction kinetics requires two key ingredients: the shape of a suitable 'free-energy surface' that describes the

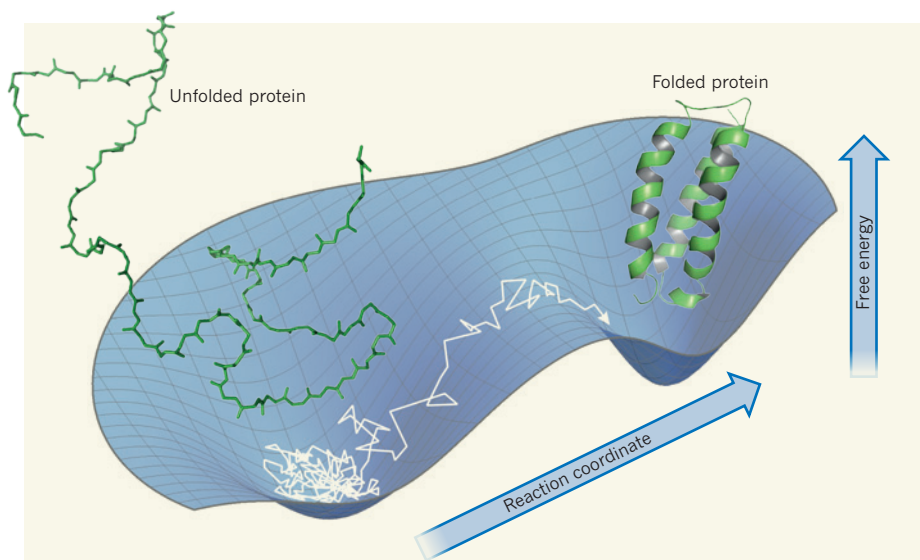


Figure 1 | Barrier crossing in protein folding. Many molecular processes can be described in terms of the diffusion of a particle on a free-energy surface, which depicts how the combined effects of energy and entropy change along a suitably chosen coordinate that represents the progress of a reaction. Here, a protein in its unfolded state corresponds to a basin on a free-energy surface; the protein must cross a free-energy barrier to reach its folded state, which constitutes another basin. The white arrow indicates the diffusive passage of the protein across the surface. Chung and Eaton¹ have used optical single-molecule experiments to probe the dynamics of the process at the top of the barrier.

*This article and the paper under discussion¹ were published online on 23 October 2013.

(in which ensembles of molecules are heated by irradiation with a laser) for very small barriers and, more recently, by single-molecule measurements^{5,9}.

What has remained a daunting challenge, however, is quantifying the key ingredients for a Kramers-like description of protein-folding reactions, especially the role of internal friction and how it changes as folding proceeds. Earlier work has shown that internal friction can be a critical factor in folding kinetics^{10,11}; that it can be highly localized to specific regions of the free-energy surface¹²; and that its contribution tends to increase as unfolded proteins become more compact¹³. Now, Chung and Eaton have investigated the nature of the barrier for the folding of individual molecules directly.

The authors measured the transition-path times for a protein (called α_3D) as a function of temperature and solvent viscosity to reveal characteristic signatures of both the solvent and internal friction. The presence of a multitude of simultaneous inter- and intramolecular interactions that slow down barrier crossing might explain why Kramers' theory applies in this case. By contrast, for some reactions of small molecules, barrier crossing can be so rapid that the solvent cannot keep up, and the simple theory fails³.

Chung and Eaton's results allowed them to estimate the height of the barrier directly — a difficult task in general, because of the large entropic contributions to the folding process, but a fundamental one, because the barrier height is a key determinant of kinetics. To model the shape of the free-energy surface, the authors took advantage of improvements in computational methods that allow simulations of protein folding in atomic detail¹⁴ and that agree remarkably well with experimental folding rates and transition-path times.

Two goals, however, have yet to be achieved: resolving the sequence of events that occur on the top of the barrier directly from single-molecule experiments, rather than from simulations; and understanding the molecular origin of internal friction. It is still unclear whether internal friction is dominated by steric hindrance (clashes of chemical groups) during rotations about the bonds in the polypeptide chain, by the transient formation of intramolecular hydrogen bonds or of clusters of hydrophobic groups, or by other short-lived interactions that must be broken to allow correct folding to proceed^{10,12}. However, the convergence of results from sophisticated experiments such as those reported by Chung and Eaton and results from simulations is a promising development, because it will increase our understanding of the detailed mechanisms of biological dynamics at the molecular scale. That will allow us to identify the requirements for applying Kramers' theory, which is widely used for describing dynamic processes in physics and chemistry³. ■

Benjamin Schuler is in the Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland. **Jane Clarke** is a Wellcome Trust Senior Research Fellow in the Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. e-mails: schuler@bioc.uzh.ch; jc162@cam.ac.uk

1. Chung, H. S. & Eaton, W. A. *Nature* **502**, 685–688 (2013).
2. Kramers, H. A. *Physica* **7**, 284–304 (1940).
3. Hänggi, P., Talkner, P. & Borkovec, M. *Rev. Mod. Phys.* **62**, 251–341 (1990).
4. Bryngelson, J. D. & Wolynes, P. G. *J. Phys. Chem.* **93**, 6902–6915 (1989).
5. Chung, H. S., McHale, K., Louis, J. M. & Eaton, W. A.

- Science* **335**, 981–984 (2012).
6. Gopich, I. V. & Szabo, A. J. *Phys. Chem. B* **113**, 10965–10973 (2009).
7. Matouschek, A., Kellis, J. T. Jr, Serrano, L. & Fersht, A. R. *Nature* **340**, 122–126 (1989).
8. Yang, W. Y. & Gruebele, M. *Nature* **423**, 193–197 (2003).
9. Yu, H. et al. *Proc. Natl Acad. Sci. USA* **109**, 14452–14457 (2012).
10. Wensley, B. G. et al. *Nature* **463**, 685–688 (2010).
11. Hagen, S. J. *Curr. Protein Peptide Sci.* **11**, 385–395 (2010).
12. Borgia, A. et al. *Nature Commun.* **3**, 1195 (2012).
13. Soranno, A. et al. *Proc. Natl Acad. Sci. USA* **109**, 17800–17806 (2012).
14. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. *Science* **334**, 517–520 (2011).

WATER MANAGEMENT

The data gap

A comprehensive search identifies a global dearth of data on the generation, treatment and use of wastewater. Remedying this situation will help policy-makers to better legislate for the management of this precious resource.

BLANCA JIMÉNEZ CISNEROS

An interesting but worrying analysis of the situation of wastewater assessment around the world is presented by Sato et al.¹ in *Agricultural Water Management*. In a thorough review of different sources of information, including scientific literature and government reports, from a spectrum of countries representing different regions and socio-economic conditions, the authors reveal that data on the generation, treatment and use of wastewater around the world are scarce and poorly reported.

Of the 181 countries assessed in the study, Sato and colleagues found that data were available on all three aspects of wastewater for only 55, and that no information at all was available from 57 countries. Only 37% of the available data were reported in the period 2008–12. Although many water-management experts are already well aware of this state of affairs, the authors' findings demonstrate to a broader audience how policy-makers around the world have, to a large extent, failed to implement policies to gather data on wastewater. This is disappointing, because such information is crucial to making informed decisions on several issues, including pollution management and how the use of wastewater could contribute to addressing water-scarcity issues and the potential threats of climate change in some regions. Water, in contrast to many other natural resources, is renewable, but in an era in which the recycling of solid waste and used materials is strongly promoted in greener societies, the lack of discussion about wastewater generation,

and especially its reuse, is remarkable.

It is worth considering some possible 'invisible' explanations for the lack of data. One is an inherent rejection of wastewater, owing to its association with water-borne diseases. Another is that, unlike the clean water that users have to pay for, wastewater may be considered to be without value, meaning that there is little interest in measuring what happens to it. But in a world in which the demand for this finite resource is increasing, wastewater — or used water, as I prefer to call it — should also be given a value and be measured. This is starting to happen in parts of the world where water is scarce and farmers, for example, must pay for wastewater.

The use of untreated wastewater for agriculture is a public-health concern.

Indeed, although many policy-makers have yet to adequately consider the potential of used water, the opposite is true of many farmers (Fig. 1). Both treated and untreated wastewater is already used for irrigation in many regions, and this form of water reuse is increasing, although Sato and colleagues' study fails to catch this trend because it is based on data not older than 2000. By reusing water instead of relying on rain-fed irrigation, farmers may be able to sow three or four crops per year, rather than only one or two. Used water also contains nutrients, such as nitrogen, phosphorus and organic matter, that enrich the soil and increase crop yields. In fact, the presence of such usable compounds in wastewater may prove to be a motive for



Figure 1 | Renewable resource. Wastewater is increasingly being used for irrigation purposes, but a lack of data on wastewater generation, treatment and use¹ is hampering the development of policies regarding this resource.

enhancing water-reuse practices. Natural reserves of phosphorus are rapidly declining, and the phosphorus industry has recently called for action to save and recycle this fertilizer². Another advantage of the reuse of water for irrigation is that it contributes to the recharging of aquifers, and thereby becomes a new source of usable water.

Thus, the attractions of using wastewater,

particularly for farmers in low-income regions, are obvious. But in many cases water is being reused without proper policies and practices. Sato *et al.* find that high-income countries, on average, treat 70% of their wastewater, but this figure drops to 8% in low-income countries, and the reuse of untreated wastewater has been estimated³ to be about 5–8 times greater than that of treated used water. The use of untreated

wastewater for agriculture is a public-health concern: wastewater can function as a vector for diarrhoeal diseases when polluted water is ingested with crops that are consumed raw.

Although irrigation is a primary focus of water-reuse issues, wastewater-management considerations should not be confined to agricultural areas. Cities may make up tiny areas of our planet, but they demand large quantities of water and food, and produce a vast amount of used water. Much of this water — and the nutrients it contains — could be reclaimed to help to produce food for urban dwellers. Methods to treat water are becoming available that cost less than conventional methods and also provide the ability to recycle nutrients⁴. Thus, researchers are developing the means to establish site-specific and cost-effective approaches to water reuse, but the data gap must be closed before such policies can be effectively designed. ■

Blanca Jiménez Cisneros is in the Division of Water Sciences, International Hydrological Programme, UNESCO, 75732 Paris Cedex 15, France.

e-mail: bjimenez-cisneros@unesco.org

1. Sato, T., Qadir, M., Yamamoto, S., Endo, T. & Zahoor, A. *Agric. Water Mgmt* **130**, 1–13 (2013).
2. Cordell, D., Drangert, J.-O. & White, S. *Glob. Environ. Change* **19**, 292–305 (2009).
3. Jiménez, B. & Asano, T. (eds) in *Water Reuse: An International Survey of Current Practice, Issues and Needs* 3–26 (IWA, 2008).
4. Jiménez, B., Mara, D., Carr, R. & Brissaud, F. in *Wastewater Irrigation and Health: Assessing and Mitigating Risk in Low-income Countries* (eds Dreschel, P. *et al.*) 149–170 (Earthscan, 2010).

DMSP by plankton and other marine microorganisms. Raina *et al.* studied the larvae of *Acropora* corals, which lack algal symbionts when first spawned. Keeping the larvae in alga-free conditions, the researchers found that the DMSP content of the corals increased by up to 54% over time as the larvae settled and matured, suggesting that the juvenile corals were producing it themselves.

The authors then dug deep into the genomes of *Acropora millepora* and *Acropora digitifera*, looking for evolutionary evidence of DMSP synthesis. They found that two genes encoding enzymes known to be involved in DMSP synthesis in diatoms — NADPH reductase and AdoMet-dependent methyltransferase — have clear orthologues in both *Acropora* and *Symbiodinium*. This suggests that the function of these enzymes is evolutionarily conserved between diatoms, alveolates (protist organisms that include the dinoflagellates), green plants and corals.

Raina and colleagues also analysed gene-transcript levels in *A. millepora* and found that the gene encoding NADPH reductase was highly expressed throughout all the coral's life stages. Expression of the gene

MARINE BIOLOGY

Coral animals combat stress with sulphur

Photosynthetic algal symbionts of corals produce sulphur substances that are involved in the regulation of ocean temperatures. In a twist to the tale, it emerges that coral animals produce the same compounds. SEE LETTER P.677

GRAHAM JONES

Coral reefs owe their success to a symbiosis between the animal host (the coral polyp) and intracellular photosynthetic dinoflagellate algae of the genus *Symbiodinium*, which supply up to 95% of the host's energy requirements. *Symbiodinium* produce abundant amounts of the sulphur compound dimethylsulphoniopropionate (DMSP) and the volatile trace sulphur gas dimethylsulphide (DMS)^{1–3} — substances that have been proposed⁴ to be involved in a climate-feedback

cycle. On page 677 of this issue, Raina *et al.*⁵ use chemical, genomic and molecular approaches to reveal that coral polyps also produce DMSP, in the absence of their algal symbionts. This biosynthesis may help the corals to survive conditions of thermal stress, and it adds the coral polyp as a possible player in climate regulation in areas rich in reefs*.

Reef-building corals, such as *Acropora* species, are prolific producers of DMSP and DMS. DMS is also formed during the breakdown of

*This article and the paper under discussion⁵ were published online on 23 October 2013.

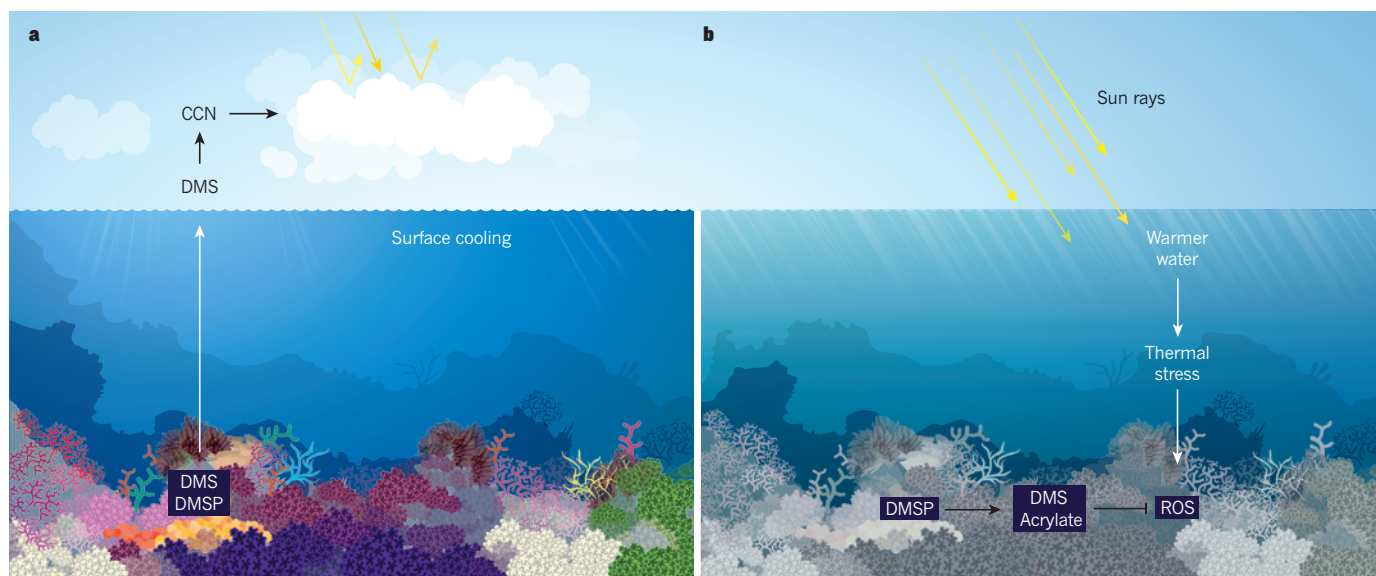


Figure 1 | Corals and DMSP. **a**, The symbiotic algae of many corals produce dimethylsulphoniopropionate (DMSP) and the gas dimethylsulphide (DMS). When DMS enters the atmosphere above reefs, it becomes oxidized to form sulphate aerosols, which act as cloud-condensation nuclei (CCN), around which water vapour condenses to form low-level clouds, which cool surface waters. Raina *et al.*⁵ show that coral host animals also produce substantial quantities of DMSP. **b**, The authors propose that DMSP synthesis by the host helps corals to survive conditions of thermal stress, which can lead to bleaching (the loss of algal symbionts) and the accumulation of toxic reactive oxygen species (ROS). Two breakdown products of DMSP, DMS and acrylate, are effective scavengers of ROS.

encoding the methyltransferase was high in the early stages, but decreased after the larvae settled, and remained relatively low in adult corals. NADPH reductase is used in other cellular pathways in addition to that leading to DMSP production⁶, which could explain why the expression levels of this enzyme are so high. By contrast, the methyltransferase is unique to the DMSP-production pathway^{6,7}. Its high level of expression in early life stages (before the corals acquire their symbionts) is probably the reason why the authors detected such high concentrations of DMSP in coral juveniles. However, the observation that reductase levels decrease when the coral acquires *Symbiodinium* suggests that the large amounts of DMSP produced by the algae influence the DMSP production by the host.

When seawater temperatures and light levels are high, the symbiosis between a coral polyp and its algae can break down, such that the algae are expelled. This process, known as coral bleaching, is predicted to increase as a result of global warming. When Raina *et al.* subjected the alga-free juvenile corals to thermal stress of 32 °C, they found that their DMSP levels increased by up to 76% compared with unstressed animals at 27 °C. By contrast, concentrations of a DMSP-breakdown product, acrylate, decreased. Both acrylate and DMS are extremely efficient scavengers of hydroxyl radicals and other reactive oxygen species (ROS) in phytoplankton⁸, and levels of ROS increase during thermal stress. These results add weight to existing evidence that DMSP and its breakdown products are involved in detoxification of ROS during thermal stress in corals⁹. Furthermore, Raina *et al.* suggest that production of DMSP by

polyps might help the coral to respond to thermal stress even when those conditions have resulted in the expulsion of the algal symbionts (Fig. 1).

Raina *et al.* conclude that corals are likely to be the exception, rather than the rule, in regard to DMSP production by marine invertebrates harbouring photosynthetic symbionts. The genus *Symbiodinium* is highly diverse, containing nine distinct clades and subclades that display varying thermal tolerance. Corals that experience bleaching induced by thermal stress usually display a subsequent increase in the concentration of *Symbiodinium* from clades with greater thermal tolerance¹⁰. It is possible that DMSP production by the host animal gives bleached corals a survival period during which more thermally tolerant symbionts can be acquired. This suggests that past periods of climate change may have led to the acquisition of this adaptive mechanism for overcoming stress.

These findings not only help us to understand how corals may respond to changing ocean temperatures, but may also contribute to our understanding of climate-feedback cycles. According to the CLAW hypothesis⁴, atmospheric oxidation of DMS produced by marine organisms will result in the generation of sulphate aerosols, which are the precursors of cloud-condensation nuclei. Because these nuclei attract water vapour, they can form low-level clouds, which will decrease solar radiation and thereby reduce sea surface temperatures (Fig. 1). Although this hypothesis has been disputed¹¹, there is evidence^{12–16} that enhanced levels of low cloud, and sea surface temperature regulation, occur in areas with high coral-reef biomass, such as

the western Pacific. Thus, it seems plausible that reefs contribute to local climate regulation through DMSP and DMS production by both coral polyps and their symbionts. But the thermal tolerance of the coral symbiosis goes only so far, and if this is exceeded and coral cover declines — as is being seen in response to current warming trends — this regional feedback loop may shut down. ■

Graham Jones is in the Marine Ecology Research Centre, Southern Cross University, Lismore, New South Wales 2480, Australia. e-mail: graham.jones@scu.edu.au

1. Broadbent, A. D., Jones, G. B. & Jones, R. A. *Estuar. Coast. Shelf Sci.* **55**, 547–555 (2002).
2. Van Alstyne, K. L., Dominique, V. J. III & Muller-Parker, G. *Coral Reefs* **28**, 167–176 (2009).
3. Yost, D. M. & Mitchelmore, C. L. *J. Exp. Mar. Biol. Ecol.* **395**, 72–79 (2010).
4. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
5. Raina, J.-B. *et al.* *Nature* **502**, 677–680 (2013).
6. Stefels, J. J. *Sea Res.* **43**, 183–197 (2000).
7. Gage, D. A. *et al.* *Nature* **387**, 891–894 (1997).
8. Sunda, W., Kieber, D. J., Kiene, R. P. & Huntsman, S. *Nature* **418**, 317–320 (2002).
9. Jones, G. B. *et al.* *Environ. Chem.* **4**, 310–322 (2007).
10. Jones, A. M. *et al.* *Proc. R. Soc. B* **275**, 1359–1365 (2008).
11. Quinn, P. K. & Bates, T. S. *Nature* **480**, 51–56 (2011).
12. Modini, R. L. *et al.* *Atmos. Chem. Phys.* **9**, 7607–7621 (2009).
13. Takahashi, Y. *et al.* *Atmos. Chem. Phys.* **10**, 1577–1584 (2010).
14. Kleypas, J. M., Danabasoglu, G. & Lough, J. M. *Geophys. Res. Lett.* **35**, L03613 (2008).
15. Leahy, S. M., Kingsford, M. J. & Steinberg, C. R. *PLoS ONE* **8**, e70400 (2013).
16. Fischer, E. & Jones, G. B. *Biogeochemistry* **110**, 31–46 (2012).

Arteriolar niches maintain haematopoietic stem cell quiescence

Yuya Kunisaki^{1,2}, Ingmar Bruns^{1,2,3*}, Christoph Scheiermann^{1,2*†}, Jalal Ahmed^{1,4*}, Sandra Pinho^{1,2}, Dachuan Zhang^{1,2}, Toshihide Mizoguchi^{1,2}, Qiaozhi Wei^{1,2}, Daniel Lucas^{1,5}, Keisuke Ito^{1,2,5}, Jessica C. Mar^{6,7}, Aviv Bergman⁶ & Paul S. Frenette^{1,2,5}

Cell cycle quiescence is a critical feature contributing to haematopoietic stem cell (HSC) maintenance. Although various candidate stromal cells have been identified as potential HSC niches, the spatial localization of quiescent HSCs in the bone marrow remains unclear. Here, using a novel approach that combines whole-mount confocal immunofluorescence imaging techniques and computational modelling to analyse significant three-dimensional associations in the mouse bone marrow among vascular structures, stromal cells and HSCs, we show that quiescent HSCs associate specifically with small arterioles that are preferentially found in endosteal bone marrow. These arterioles are ensheathed exclusively by rare NG2 (also known as CSPG4)⁺ pericytes, distinct from sinusoid-associated leptin receptor (LEPR)⁺ cells. Pharmacological or genetic activation of the HSC cell cycle alters the distribution of HSCs from NG2⁺ periarteriolar niches to LEPR⁺ perisinusoidal niches. Conditional depletion of NG2⁺ cells induces HSC cycling and reduces functional long-term repopulating HSCs in the bone marrow. These results thus indicate that arteriolar niches are indispensable for maintaining HSC quiescence.

Somatic stem cells self-renew to maintain tissue homeostasis for the lifetime of organisms through tightly controlled proliferation and differentiation^{1–3}. In the bone marrow, recent studies have highlighted the critical influence of the microenvironment in regulating HSCs⁴. Deletion of genes involved in maintaining cell cycle quiescence has shown that unchecked HSC proliferation often leads to stem cell exhaustion^{5–8}. Although most HSCs are quiescent under homeostasis⁹, they can undergo activation, for example by interferon-mediated signals^{7,10,11}. This raises the question of whether quiescent and proliferative HSCs are found in the same niche.

The identification of cellular constituents of the HSC niche has recently been the subject of intense study. Initial reports have suggested that osteoblasts are niche cells as HSCs tend to localize near endosteal surfaces¹², and that factors increasing osteoblast numbers can also increase the number of HSCs^{12,13}. N-cadherin⁺ osteoblasts have been proposed to promote HSC quiescence through direct contact^{12,14} and secretion of angiopoietin 1 (ref. 15) or osteopontin^{16,17}. However, the synthesis of these factors is not specific to osteoblasts, and other studies have found that most bone marrow HSCs are found near sinusoidal endothelial cells¹⁸ and perivascular stromal cells including CXCL12-abundant reticular cells^{19,20}, Nestin⁺ mesenchymal stem cells²¹ or LEPR⁺ cells²². On the basis of these data, a prevalent unifying interpretation of the literature has been that the osteoblastic and vascular niches confer distinct microenvironments promoting quiescence and proliferation, respectively^{2,23}. However, this popular concept has not been supported by rigorous analyses. To evaluate this issue, we have used novel three-dimensional bone marrow imaging combined with computational modelling to assess meaningful relationships between endogenous quiescent HSCs and stromal structures. These studies have allowed us to identify distinct vascular niches mediating stem cell quiescence and proliferation.

HSCs associate with bone marrow arterioles

To gain detailed insight into the three-dimensional structure of the HSC niche, we prepared whole-mount tissues to visualize by confocal immunofluorescence imaging the architecture of long-bone and sternal marrow over ~75 μm thickness (Fig. 1a, b and Extended Data Fig. 1a, b). To specifically label bone marrow endothelial cells, we performed *in vivo* staining (Extended Data Fig. 1c–e). Whole-mount assessment of the femoral bone marrow vasculature revealed an even distribution of the sinusoidal network that occupies $30 \pm 5\%$ of the bone marrow volume (Fig. 1c, d) and where individual sinusoidal vessels are regularly spaced by $46 \pm 1 \mu\text{m}$ (Extended Data Fig. 1f). In addition to the sinusoidal network, three-dimensional visualization of the bone marrow vasculature highlighted the presence of small-calibre (10–20 μm) Sca-1 (stem cell antigen-1)^{hi} VEGFR2 (also known as KDR or FLK1)⁺ VEGFR3 (also known as FLT4)⁺ arterioles²⁴, which were found predominantly in close proximity to the bone²⁵ and comprised a much smaller volumetric fraction, $1.2 \pm 0.1\%$ of the bone marrow (Fig. 1a–d and Extended Data Fig. 2a, b). The vessels were confirmed as bona fide arterioles by their pronounced TIE2 (also known as TEK)–green fluorescent protein (GFP) expression²⁶, absence of *in vivo* staining with the sinusoid-specific Dil-labelled acetylated low-density lipoprotein (Dil-Ac-LDL)²⁶, and strong staining with the artery-specific dye Alexa Fluor 633 (ref. 27) (Fig. 1e and Extended Data Fig. 2b–f). The distribution of phenotypic CD150⁺ CD48[–] CD41[–] Lineage[–] HSCs¹⁸ was not uniform as they localized predominantly to the peripheral zone rather than in close proximity to the central vein in the long-bone bone marrow (Fig. 1d and Extended Data Fig. 3a). We validated the identification of rare phenotypic HSCs by using whole-mount preparations of the mouse sternum²⁸. Unlike long bones, which are mostly occupied by adipocytes in adult humans²⁹, the sternum exhibits rich haematopoietic activity in both species. The mouse sternum is made of six thin bone

¹Ruth L. and David S. Gottesman Institute for Stem Cell and Regenerative Medicine Research, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ²Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ³Department of Hematology, Oncology and Clinical Immunology, Heinrich-Heine-University, Düsseldorf 40225, Germany. ⁴Mount Sinai School of Medicine, New York, New York 10029, USA. ⁵Department of Medicine, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ⁶Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ⁷Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [†]Present address: Walter-Brendel-Centre of Experimental Medicine, Ludwig-Maximilians-Universität, München, Munich 81377, Germany.

*These authors contributed equally to this work.

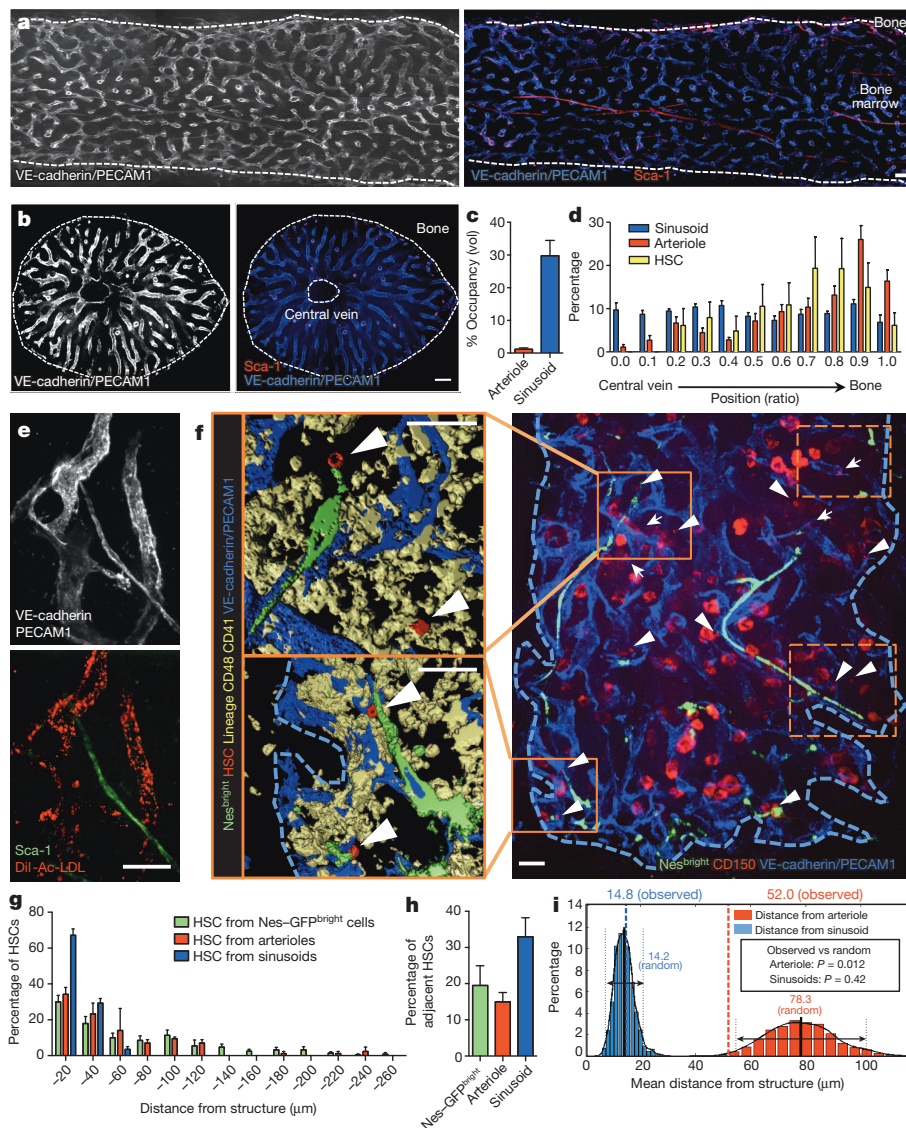


Figure 1 | Spatial relationships between HSCs and the bone marrow vasculature.

a, **b**, Longitudinally (**a**) and transverse-shaved (**b**) whole-mount images of the mouse femoral bone marrow stained with anti-PECAM1, anti-vascular endothelial (VE)-cadherin and anti-Sca-1 antibodies. Scale bar, 100 μ m. **c**, Bone marrow volumes occupied by arterioles or sinusoids. $n = 6$ areas from 3 mice. **d**, Distribution of sinusoids, arterioles and HSCs in the femoral bone marrow. $n = 6$ mice. **e**, Whole-mount images of sternum stained with anti-VE-cadherin, anti-PECAM1, anti-Sca-1 antibodies and Dil-Ac-LDL. Scale bar, 50 μ m. **f**, Illustrative example of whole-mount sternal bone marrow with three-dimensionally reconstructed images. Dashed squares are shown in Extended Data Fig. 3b. Arrowheads denote HSCs, arrows show CD150⁺ Lin/CD48/CD41⁺ cells. Scale bars, 50 μ m. **g**, **h**, Distances between HSCs and Nes-GFP^{bright} cells ($n = 98$ HSCs from 5 mice), arterioles or sinusoids ($n = 119$ HSCs from 5 mice) (**g**) and percentages of adjacent HSCs (distance = 0) (**h**) in the sternal bone marrow. In **g**, HSC distances from structures were binned into 20- μ m intervals. **i**, Probability distributions of mean distances from simulations of randomly positioned HSCs on maps of sternal bone marrow in relation to sinusoids (blue) or arterioles (red). Mean distances observed *in situ* (dashed line) are shown in relation to the grand mean \pm 2 s.d. (solid and dotted lines, respectively).

compartments in which all HSCs can be enumerated (Fig. 1f and Extended Data Fig. 3b, c).

The absolute numbers of HSCs per sternal bone compartment, determined by whole-mount imaging, was similar to assessment by fluorescence-activated cell sorting (FACS), indicating that the staining of endogenous HSCs was reliable (Extended Data Fig. 3d). We next examined the spatial relationships between individual HSCs and arterioles and sinusoids by whole-mount immunostaining of sternal bone marrow. We found that $36.8 \pm 1.8\%$ and $67.1 \pm 3.5\%$ of HSCs were located within 20- μ m distance from arterioles and sinusoids, and $14.9 \pm 2.5\%$ and $32.9 \pm 5.2\%$ were adjacent to those vessels, respectively (Fig. 1g, h and Extended Data Fig. 3e, f). The high proportion of HSCs near sinusoids is consistent with previous reports^{18,19}.

As the bone marrow sinusoidal network was dense and appeared to be regularly interspaced, we evaluated in more detail whether the spatial distribution of HSCs was independent of the sinusoidal network or significantly biased in relation to a putative sinusoidal niche. To address this issue, we developed a computational simulation to assess the statistical relevance of these associations (Extended Data Fig. 4a–c). We carried out 1,000 simulations in which 20 HSCs were randomly positioned on images of whole-mount-prepared sterna. We found that the mean distance observed *in situ* to sinusoids could not be statistically differentiated from that of randomly placed HSCs (14.2 μ m and 14.8 μ m for random and observed, respectively; Fig. 1i).

By contrast, the observed mean distance to arterioles (52.0 μ m) was highly statistically different from that of randomly placed HSCs (78.3 μ m, $P = 0.012$; Fig. 1i). Thus, these data suggest that associations between HSCs and arterioles are highly significant whereas the close relationship observed with sinusoids could not be discriminated from that of a random distribution.

Distinct Nestin⁺ cells segregate with distinct vessels

Nestin⁺ perivascular cells have been identified as HSC niche cells containing all bone marrow mesenchymal stem cell activity²¹. Detailed three-dimensional imaging and FACS analyses revealed two distinct types of Nestin-GFP (hereafter Nes-GFP)⁺ cells in the bone marrow on the basis of their GFP expression levels and cellular morphology (Fig. 2a, b and Extended Data Fig. 5a). Nes-GFP^{bright} cells were much rarer ($\sim 0.002\%$ of bone marrow) than Nes-GFP^{dim} cells (Fig. 2a and Extended Data Fig. 5b), and found exclusively along arterioles in both sternal and long-bone bone marrow (Extended Data Figs 2d, e and 5c–h). Consistent with this observation, similar HSC distribution patterns were obtained when plotting distances from arterioles or from Nes-GFP^{bright} cells (hereafter referred to as periarteriolar Nestin cells, Nes^{peri}; Fig. 1g, h and Extended Data Fig. 3e, f, $P = 0.97$). The more abundant Nes-GFP^{dim} population was reticular in shape (hereafter referred to as Nes^{retic} cells) and was largely associated with sinusoids (Fig. 2b and Extended Data Figs 3e, f and 5g). Nes^{peri} cells, but not

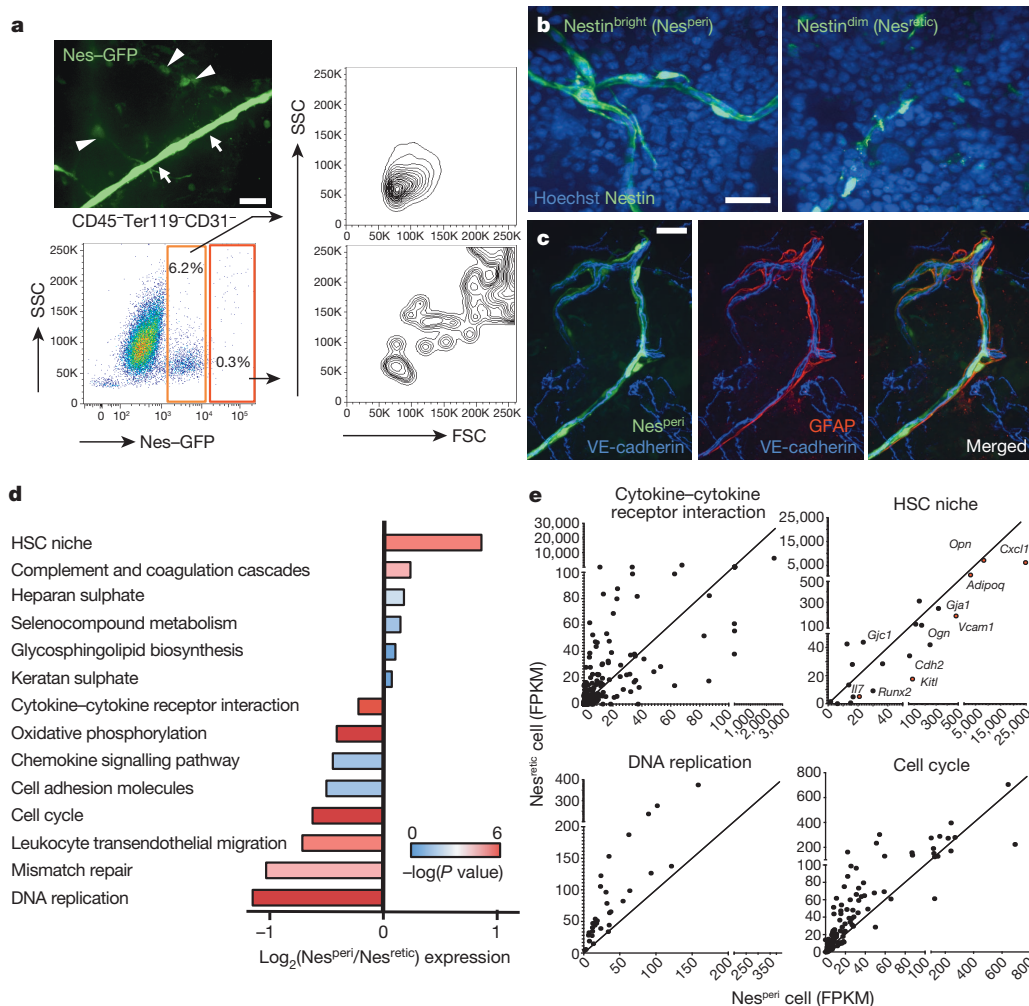


Figure 2 | Nestin⁺ cell subsets define distinct vascular structures. **a**, A whole-mount image of sternal bone marrow with Nes-GFP^{bright} (arrows) and Nes-GFP^{dim} cells (arrowheads) and FACS analysis of bone marrow stromal cells from Nes-GFP mice. Data were reproduced in at least 5 mice. FSC, forward scatter; SSC, side scatter. **b, c**, Whole-mount images of sternum from Nes-GFP mice stained with anti-GFP and Hoechst 33342 (b) or anti-glial fibrillary acidic protein (GFAP) antibody (c). **d, e**, RNA-seq analysis of Nes^{retic} and Nes^{peri} cells. **d**, Differentially expressed pathways. Mean of log fold change of Nes^{peri}/Nes^{retic} expression was plotted for each pathway with colour-coded P values. **e**, Dot plots showing gene expression levels in fragments per kilobase of exon per million fragments mapped (FPKM) for representative pathways and functional groups.

Nes^{retic} cells, were decorated with tyrosine hydroxylase⁺ sympathetic nerves and glial fibrillary acidic protein⁺ Schwann cells, which have been reported to modulate HSC behaviour in the bone marrow^{30–32} (Fig. 2c and Extended Data Fig. 5i). Although both Nestin⁺ cell subsets accounted for the mesenchymal progenitor activity (fibroblastic colony-forming units; CFU-F) in bone marrow, most CFU-F were contained in Nes^{peri} cells (Extended Data Fig. 5j). To get additional insight into potential differences between these two stromal subsets, we subjected Nes^{peri} and Nes^{retic} cells to RNA sequencing (RNA-seq) analyses. Comparison of the whole-genome transcriptome revealed differential enrichment of several molecular pathways and functional clusters, including HSC niche, cytokine–receptor interactions, DNA replication and cell cycle (Fig. 2d, e and Supplementary Table 1). Indeed, Nes^{peri} cells showed significantly higher expression of genes associated with the HSC niche (Fig. 2d, e and Extended Data Fig. 5k). In addition, expression of genes in two pathways, DNA replication and cell cycle, was significantly enriched in Nes^{retic} cells (Fig. 2d, e), suggesting that the two subsets had differing cell cycle characteristics.

Arteriolar Nes^{peri} niche cells are quiescent

On the basis of these results, we evaluated in more detail the cell cycle status of bone marrow stromal cells, and more specifically of Nes^{retic} and Nes^{peri} cells. Expression of the nuclear proliferation markers Ki-67 and proliferation cell nuclear antigen (PCNA) in Nes^{peri} cells was significantly lower than in Nes^{retic} cells and CD45⁺ Ter119⁺ CD31⁺ Nes-GFP⁺ stromal cells (Fig. 3a, b), which was confirmed by FACS analysis (Fig. 3c, d). To test the relevance of niche cell quiescence, we challenged mice with 5-fluorouracil (5-FU), a drug that kills cycling

cells. After 5-FU administration, Nes^{peri} cells were largely preserved structurally and numerically compared to Nes^{retic} cells (Fig. 3e–g), confirming their quiescent state.

Dormant HSCs associate with arterioles

We proposed that the arteriolar quiescent niche cells may themselves harbour quiescent HSCs. Consistent with previous studies^{10,33}, we found that most HSCs were not actively cycling, but ~20% of HSCs were in the non-quiescent G1 phase of the cell cycle (Extended Data Fig. 6a). Prolonged quiescence has been referred to as dormancy and tracked as label-retaining cells¹⁰. We first evaluated the spatial associations of dormant HSCs with Nes^{peri} cells by label-retaining cell pulse-chase experiments with 5-ethynyl-2'-deoxyuridine (EdU) in the long-bone bone marrow. Significantly more EdU⁺ CD150⁺ CD48⁺ CD41⁺ Lineage⁺ HSCs compared to total HSCs were located within a 20-μm distance from Nes^{peri} cells (Fig. 4a, b and Extended Data Fig. 6b; 65.8 ± 0.8% versus 44.3 ± 1.8%, *P* = 0.016). We next analysed the cell cycle status and distribution of HSCs in relation to arterioles with Ki-67 staining in the sternal bone marrow and found that a significantly higher proportion of Ki-67⁺ quiescent HSCs were located within a 20-μm distance compared to Ki-67⁺ non-quiescent HSCs (Fig. 4c, d, g and Extended Data Fig. 6c, d). There was no significant difference in the proportions of adjacent cells between quiescent and non-quiescent HSCs (Extended Data Fig. 6b, c), suggesting that proximity to arterioles, rather than direct contact, promotes HSC quiescence. As osteoblasts have been suggested to be associated with quiescent HSCs^{12,14–17}, we evaluated the distribution of HSCs in relation to osteoblasts in the sternal bone marrow using transgenic

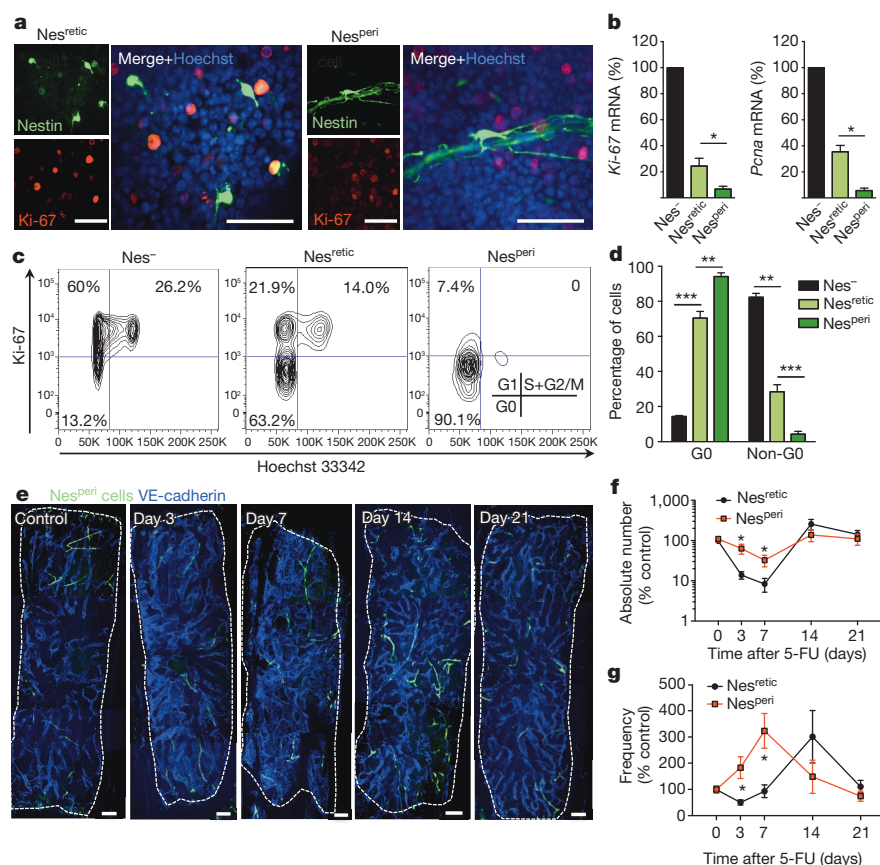


Figure 3 | Quiescent arteriolar niche cells are protected from myeloablation. **a**, Whole-mount immunostaining with anti-Ki-67 antibody of sterna from Nes-GFP mice. Scale bars, 50 μ m. **b**, Quantitative PCR analyses of Nes^{peri}, Nes^{retic} and CD45⁺ Ter119⁺ CD31⁺ Nes-GFP⁺ stromal cells. $n = 5$ independent experiments. **c, d**, Cell cycle analyses by FACS using anti-Ki-67 and Hoechst 33342 staining. Representative plots (c) and quantification (d). $n = 3$ mice. **e–g**, Whole-mount images of sternal bone marrow (e) and the kinetics of absolute numbers (f) and frequencies (g) of Nes^{peri} and Nes^{retic} cells in the long-bone bone marrow analysed by FACS after 5-FU treatment. $n = 10, 9, 9, 5, 6$ mice per time point. Scale bars, 100 μ m.

mice expressing GFP under a 2.3-kilobase fragment of the collagen promoter (Col2.3-GFP), in which osteoblasts and osteocytes express GFP³⁴. We found that $20.7 \pm 1.9\%$ of total HSCs were located within 20- μ m distance from Col2.3-GFP⁺ cells (Fig. 4e and Extended Data Fig. 6e). We performed the computational simulation with Col2.3-GFP⁺ cells and found that in contrast to arterioles, HSCs were not significantly associated with Col2.3-GFP⁺ cells (Extended Data Fig. 6f). Moreover, there was no significant difference in the distribution of quiescent and non-quiescent HSCs in relation to Col2.3-GFP⁺ cells in the sternal bone marrow (Fig. 4f, g and Extended Data Fig. 6g, $P = 0.93$). These results suggest that quiescent HSCs are preferentially associated with arterioles in both long-bone and sternal marrow.

To test this idea using functional assays, we evaluated HSC–arteriole associations in mice treated with 5-FU. We found that at nadir on days 3 and 7, the vast majority of HSCs were closely associated with Nes^{peri} cells (control/day 3/day 7; $34.4 \pm 1.0/71.1 \pm 3.9/70.6 \pm 4.0\%$ in 0–20 μ m, respectively) and that HSCs expanded in further distance from Nes^{peri} cells upon recovery in the sternal bone marrow (Fig. 4h and Extended Data Fig. 6h, i). These data indicate that arterioles indeed provide a milieu promoting quiescence for both HSCs and their niche cells, and a safe harbour to protect them against genotoxic insults.

We next tested whether changes in HSC cycling would lead to altered distribution. To this end, we injected mice with polyinosinic-polycytidylic acid (Poly(I:C)), a compound reported to induce proliferation of HSCs via interferon- α production, which can lead to HSC exhaustion^{7,8}. Poly(I:C) treatment increased the fraction of non-quiescent HSCs (Extended Data Fig. 7a) and altered their distribution in relation to arterioles in that the percentage of HSCs within 20 μ m was significantly reduced compared to phosphate-buffered saline-treated animals in the sternal bone marrow (Extended Data Fig. 7b, $P = 0.007$). Furthermore, we evaluated a genetic mouse model deficient in promyelocytic leukaemia protein (*Pml*^{−/−}) that was shown to compromise HSC quiescence in a cell-autonomous manner and caused HSC exhaustion⁶.

Remarkably, endogenous HSCs were distributed significantly further away from arterioles in *Pml*^{−/−} mice compared to wild-type animals (Extended Data Fig. 7c, $P = 1.9 \times 10^{-6}$). Notably, the location of quiescent Ki-67[−] cells was markedly affected (Extended Data Fig. 7d, e). These data suggest the possibility that there are spatially distinct niches for quiescent and proliferating HSCs in the bone marrow.

To further confirm this notion, we treated mice with granulocyte colony-stimulating factor (G-CSF), a drug commonly used to mobilize HSCs from the bone marrow to the circulation, and that is also known to induce HSC proliferation^{35,36}. Short-term (1 or 2 days) administration of G-CSF led to altered HSC distribution further away from Nes^{peri} cells. By contrast, further treatment (4 days) led to expansion of HSCs around Nes^{peri} cells (Extended Data Fig. 7f, g). In line with these data, G-CSF also downregulated HSC retention factors such as CXCL12 in the Nestin⁺ niche cells²¹, which also may contribute to the shift from the quiescent to the proliferative (mobilizable) niches at early phase, the release of the mobilizable pool, and HSC replenishment at late phase. This model further highlights the importance of local cues in fine-tuning HSC behaviour in the bone marrow.

NG2⁺ arteriolar niche cells promote HSC quiescence

LEPR⁺ perivascular cells have been reported to represent an important source of stem cell factor (SCF) required for the maintenance of HSCs in the bone marrow²². As Nestin⁺ cells were shown to express large amounts of SCF²¹, we evaluated the overlap between Nes^{peri} or Nes^{retic} cells and LEPR⁺ cells. Using triple-transgenic *Lepr-cre/loxptdTomato/Nes-GFP* mice, we found that LEPR⁺ cells were adjacent to sinusoids in a manner similar to Nes^{retic} cells (Extended Data Fig. 8a). Indeed, Nes^{retic} cells largely overlapped (~80%) with LEPR⁺ cells whereas we found no overlap between Nes^{peri} and LEPR⁺ cells (Extended Data Fig. 8b, c). By contrast, Nes^{peri}, but not Nes^{retic}, cells were positive for the classical pericyte marker NG2 and α -smooth muscle actin³⁷ (Extended Data Fig. 9a–f). To examine further the expression of NG2 on Nes^{peri}

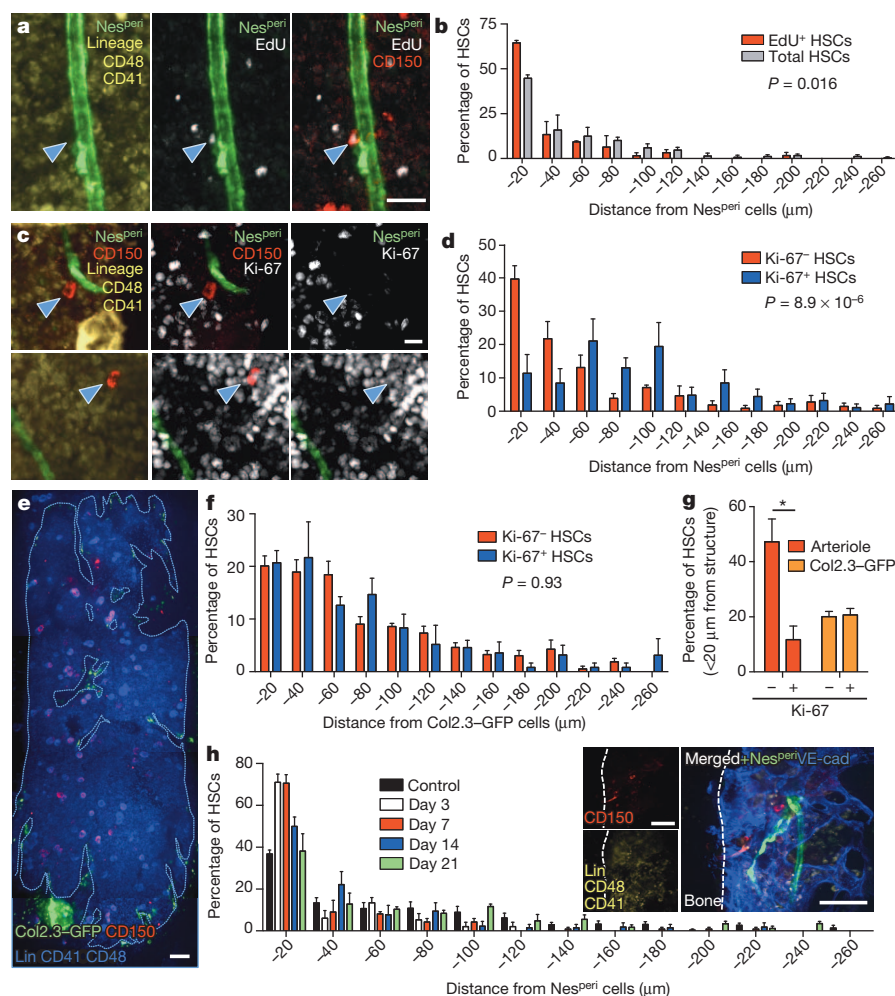


Figure 4 | Spatial relationship between arterioles and quiescent HSCs. **a, b**, Localization of dormant HSCs accessed by long-term label retention with EdU. Representative images (**a**) and distribution of total and EdU⁺ CD150⁺ CD48⁺ CD41⁻ Lineage⁻ HSCs in femoral bone marrow sections (**b**). $n = 144$, 52 HSCs from 3 mice. Two-sample Kolmogorov–Smirnov test; $P = 0.016$. Scale bar, 25 μ m. **c, d**, Localization of Ki-67⁺ and Ki-67⁻ HSCs in the sternal bone marrow. Ki-67⁺/Ki-67⁻: $39.7 \pm 3.9\%/11.4 \pm 5.6\%$ in 0–20- μ m proximity, $21.5 \pm 5.3\%/47.9 \pm 6.5\%$ > 80- μ m distance. Representative images (**c**) and distances from Nes^{peri} cells (**d**). $n = 116$, 64 HSCs from 7 mice. Two-sample Kolmogorov–Smirnov test; $P = 8.9 \times 10^{-6}$. Scale bar, 10 μ m. **e–g**, Localization of HSCs in relation to osteoblasts. **e**, Whole-mount images of sternum from Col2.3–GFP mice with HSC staining. Scale bar, 100 μ m. **f**, Distances of Ki-67⁺ and Ki-67⁻ HSCs from Col2.3–GFP cells. $n = 80$, 41 HSCs from 3 mice. Two-sample Kolmogorov–Smirnov test; $P = 0.93$. **g**, Percentage of Ki-67⁺ and Ki-67⁻ HSCs located within 20- μ m distance from arterioles ($n = 7$ mice) or Col2.3–GFP cells ($n = 3$ mice). **h**, Localization of HSCs relative to Nes^{peri} cells after 5-FU treatment. Inset shows a representative image of sternal bone marrow on day 7. $n = 98$, 39, 70, 55, 112 HSCs from 5, 9, 7, 4, 4 mice per control, day 3, day 7, day 14 and day 21 groups, respectively. Two-sample Kolmogorov–Smirnov test; day 3, $P = 2.8 \times 10^{-4}$; day 7, $P = 1.0 \times 10^{-4}$; day 14, $P = 0.0013$; day 21, $P = 0.23$ compared to control. Scale bar, 50 μ m.

cells, we generated triple-transgenic mice to label both NG2⁺ and Nestin⁺ cells (NG2 promoter/enhancer driving tamoxifen-inducible Cre; NG2-cre^{ERTM}/loxP-tdTomato/Nes-GFP mice). In these mice, 30% of Nes^{peri} cells were labelled by the NG2-cre^{ERTM} transgene (Fig. 5a, b and Extended Data Fig. 9g, h), a recombination efficiency consistent with that reported for oligodendrocyte progenitors in the brain³⁸. Notably, no labelling was observed in Nes^{retic} cells or in other stromal cells (Fig. 5b and Extended Data Fig. 9i).

To examine the function of NG2⁺ Nes^{peri} cells in HSC maintenance, we intercrossed NG2-cre^{ERTM} with an inducible diphtheria toxin receptor (iDTR) line. Tamoxifen and diphtheria toxin treatment depleted ~55% of Nes^{peri} cells without significantly affecting the vascular volume as determined by whole-mount imaging 7–16 days after diphtheria toxin administration (Fig. 5c and Extended Data Fig. 10a–c). NG2⁺ cell depletion also did not affect the expression of HSC niche genes in endothelial cells (Extended Data Fig. 10d, e). However, depletion of NG2⁺ cells altered HSC localization away from arterioles (Fig. 5d and Extended Data Fig. 10f) and switched them into non-quiescent status (Fig. 5e, f). In addition, we found that the number and frequency of HSCs in the bone marrow and spleen, as determined phenotypically (Fig. 5g and Extended Data Fig. 10g) and functionally using long-term culture-initiating cell (LTC-IC; Extended Data Fig. 10h, i) or competitive reconstitution assays (Fig. 5h), were significantly reduced 16 days after diphtheria toxin administration. These results indicate that depletion of NG2⁺ cells reduces the HSC pool rather than inducing mobilization to extramedullary sites (Extended Data Fig. 10j). Taken together, these data strongly suggest that rare NG2⁺ periaarteriolar cells promote HSC quiescence and are essential for HSC maintenance in the bone marrow.

Discussion

Here, we show that arterioles organize a critical niche that maintains HSC quiescence in the bone marrow. Arterioles are associated with both quiescent NG2⁺ niche cells and HSCs, suggesting that the vessel itself may be a critical gatekeeper of stem cell quiescence in the bone marrow. Arterioles are structurally distinct from sinusoids as they are surrounded by sympathetic nerves, layers of smooth muscle cells and matrix components to carry blood from the heart under high pressure (Supplementary Table 2). These differing characteristics translate into significant differences in the transcriptional program of arteriolar and venular endothelial cells³⁹. Likewise, arteriolar Nes^{peri} cells and sinusoidal Nes^{retic} cells can be distinguished by their individual transcriptional profiles reflecting relative enrichments in cell cycle quiescence and HSC-niche-related genes in Nes^{peri} cells. The present data show that upon severe genotoxic insult, the mitotically active Nes^{retic} sinusoidal niche cells are largely destroyed, whereas the quiescent Nes^{peri} arteriolar niche cells are chemoresistant. The association and quiescence of both haematopoietic and mesenchymal stem cells at this site may thus endow arterioles with the potential for orchestrating haematopoietic and stromal regeneration⁴⁰.

Given the recent studies showing that CXCL12 derived from mesenchymal progenitors marked by *Prx1-cre* contributes to HSC maintenance^{41,42}, it is possible that NG2⁺ Nes^{peri} cells are also targeted by *Prx1-cre*. Although Nes-GFP⁺ cells express high levels of SCF²¹, Nes-cre mediated deletion of *Scf* has no effect on HSC function²². This may be due to the low recombination efficiencies among Nes-GFP⁺ cells in this model, allowing for compensation by the non-targeted cells. In keeping with the notion that multiple niche constituents exert distinct functions, conditional deletion of *Cxcl12* by *Lepr-cre*

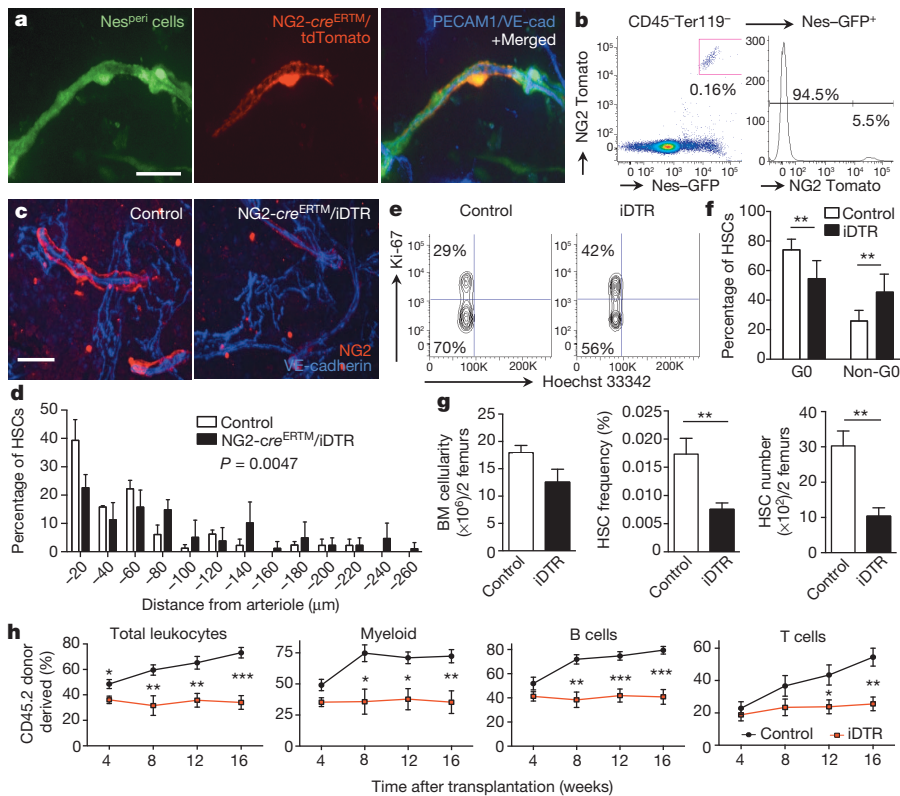


Figure 5 | NG2⁺ periaarteriolar cells are an essential constituent of the HSC niche promoting HSC quiescence. **a**, **b**, Whole-mount images of sternal bone marrow (**a**) and FACS analysis of femoral bone marrow (**b**) from NG2-*cre*^{ERTM}/loxp-tdTomato/Nes-GFP transgenic mice. Representative data from 4 mice. **c**–**h**, Analysis of HSCs after specific deletion of NG2⁺ cells using NG2-*cre*^{ERTM}/iDTR mice. **c**, Whole-mount immunostaining of sternum with anti-NG2 antibody. Representative images of 5 mice per group. **d**, HSC localization relative to arterioles in the sternal bone marrow. *n* = 69, 71 HSCs from 3, 4 mice per control and NG2-*cre*^{ERTM}/iDTR groups, respectively. Two-sample Kolmogorov–Smirnov test; *P* = 0.0047. **e**, **f**, Cell cycle analysis of HSCs using Ki-67 and Hoechst 33342 staining. Representative plots (CD150⁺ CD48⁺ Sca-1⁺ c-kit⁺ Lineage[−] gated) (**e**) and quantifications (**f**). *n* = 5 mice per group. **g**, Bone marrow (BM) cellularity, frequency and number of HSCs in the bone marrow. *n* = 6 mice per group. **h**, Quantification of long-term reconstituting HSCs by competitive reconstitution assays. *n* = 6, 9 mice per control and NG2-*cre*^{ERTM}/iDTR groups. Scale bar, 25 μm.

does not affect HSC numbers in the bone marrow, but induces their mobilization⁴².

The idea that sinusoids may represent a proliferative niche is consistent with recent studies indicating that E-selectin, an adhesion molecule constitutively expressed in certain bone marrow sinusoidal microdomains⁴³, promotes HSC proliferation and blockade of E-selectin protects HSCs following chemotherapy or γ -irradiation⁴⁴. It is therefore conceivable that a continuous exchange between the sinusoidal and arteriolar niches contributes to maintain a tightly controlled balance of the HSC pool between proliferation and dormancy.

An arteriolar niche maintaining stem cell dormancy has the potential to protect haematopoietic malignancies and metastatic cancer stem cells. Gaining knowledge of the reciprocal regulation and functional correlation of these niches will thus have important implications in understanding stem cell fate in health and cancer.

METHODS SUMMARY

B6.Cg-Tg(Cspg4-cre/Esr1*)BAK1/J, STOCK Tg (TIE2GFP) 287Sato/J (TIE2-GFP) mice, C57BL/6-Gt(ROSA)26 Sortm1(HBEGF) Awa1/J, B6.Cg-Gt(ROSA)26Sortm14(CAG-tdTomato)Hze/J, B6.129-Leprtm2(cre)Rck/J (Lepr-Cre), *Pml*^{−/−}, Nes-GFP⁴⁵ and Col2.3-GFP³⁴ mice were used in this study. Unless indicated otherwise, data are presented as mean \pm s.e.m.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 March; accepted 28 August 2013.

Published online 9 October 2013.

- Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542–545 (2010).
- Hsu, Y. C., Pasolli, H. A. & Fuchs, E. Dynamics between stem cells, niche, and progeny in the hair follicle. *Cell* **144**, 92–105 (2011).
- Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nature Rev. Genet.* **9**, 115–128 (2008).
- Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21^{cip1/waf1}. *Science* **287**, 1804–1808 (2000).

- Ito, K. *et al.* PML targeting eradicates quiescent leukaemia-initiating cells. *Nature* **453**, 1072–1078 (2008).
- Essers, M. A. *et al.* IFN α activates dormant haematopoietic stem cells *in vivo*. *Nature* **458**, 904–908 (2009).
- Sato, T. *et al.* Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon-dependent exhaustion. *Nature Med.* **15**, 696–700 (2009).
- Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *J. Exp. Med.* **202**, 1599–1611 (2005).
- Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
- Baldrige, M. T., King, K. Y., Boles, N. C., Weksberg, D. C. & Goodell, M. A. Quiescent haematopoietic stem cells are activated by IFN- γ in response to chronic infection. *Nature* **465**, 793–797 (2010).
- Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
- Sugimura, R. *et al.* Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell* **150**, 351–365 (2012).
- Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161 (2004).
- Nilsson, S. K. *et al.* Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood* **106**, 1232–1239 (2005).
- Stier, S. *et al.* Osteopontin is a hematopoietic stem cell niche component that negatively regulates stem cell pool size. *J. Exp. Med.* **201**, 1781–1791 (2005).
- Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977–988 (2006).
- Omatsu, Y. *et al.* The essential functions of adipo-osteogenic progenitors as the hematopoietic stem and progenitor cell niche. *Immunity* **33**, 387–399 (2010).
- Méndez-Ferrer, S. *et al.* Mesenchymal and hematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).
- Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
- Ema, H. & Suda, T. Two anatomically distinct niches regulate stem cell activity. *Blood* **120**, 2174–2181 (2012).
- Hooper, A. T. *et al.* Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4**, 263–274 (2009).

25. Nombela-Arrieta, C. *et al.* Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment. *Nature Cell Biol.* **15**, 533–543 (2013).
26. Li, X. M., Hu, Z., Jorgenson, M. L. & Slayton, W. B. High levels of acetylated low-density lipoprotein uptake and low tyrosine kinase with immunoglobulin and epidermal growth factor homology domains-2 (Tie2) promoter activity distinguish sinusoids from other vessel types in murine bone marrow. *Circulation* **120**, 1910–1918 (2009).
27. Shen, Z., Lu, Z., Chhatbar, P. Y., O'Herron, P. & Kara, P. An artery-specific fluorescent dye for studying neurovascular coupling. *Nature Methods* **9**, 273–276 (2012).
28. Takaku, T. *et al.* Hematopoiesis in 3 dimensions: human and murine bone marrow architecture visualized by confocal microscopy. *Blood* **116**, e41–e55 (2010).
29. Naveiras, O. *et al.* Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment. *Nature* **460**, 259–263 (2009).
30. Katayama, Y. *et al.* Signals from the sympathetic nervous system regulate hematopoietic stem cell egress from bone marrow. *Cell* **124**, 407–421 (2006).
31. Méndez-Ferrer, S., Lucas, D., Battista, M. & Frenette, P. S. Haematopoietic stem cell release is regulated by circadian oscillations. *Nature* **452**, 442–447 (2008).
32. Yamazaki, S. *et al.* Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. *Cell* **147**, 1146–1158 (2011).
33. Bhattacharya, D. *et al.* Niche recycling through division-independent egress of hematopoietic stem cells. *J. Exp. Med.* **206**, 2837–2850 (2009).
34. Kalajic, Z. *et al.* Directing the expression of a green fluorescent protein transgene in differentiated osteoblasts: comparison between rat type I collagen and rat osteocalcin promoters. *Bone* **31**, 654–660 (2002).
35. Liu, F. *et al.* *Csf3r* mutations in mice confer a strong clonal HSC advantage via activation of Stat5. *J. Clin. Invest.* **118**, 946–955 (2008).
36. Greenbaum, A. M. & Link, D. C. Mechanisms of G-CSF-mediated hematopoietic stem and progenitor mobilization. *Leukemia* **25**, 211–217 (2011).
37. Birbrair, A. *et al.* Skeletal muscle pericyte subtypes differ in their differentiation potential. *Stem Cell Res.* **10**, 67–84 (2013).
38. Zhu, X. *et al.* Age-dependent fate and lineage restriction of single NG2 cells. *Development* **138**, 745–753 (2011).
39. Chi, J. T. *et al.* Endothelial cell diversity revealed by global expression profiling. *Proc. Natl Acad. Sci. USA* **100**, 10623–10628 (2003).
40. Frenette, P. S., Pinho, S., Lucas, D. & Scheiermann, C. Mesenchymal stem cell: keystone of the hematopoietic stem cell niche and a stepping-stone for regenerative medicine. *Annu. Rev. Immunol.* **31**, 285–316 (2013).
41. Greenbaum, A. *et al.* CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* **495**, 227–230 (2013).
42. Ding, L. & Morrison, S. J. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* **495**, 231–235 (2013).
43. Sipkins, D. A. *et al.* *In vivo* imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435**, 969–973 (2005).
44. Winkler, I. G. *et al.* Vascular niche E-selectin regulates hematopoietic stem cell dormancy, self renewal and chemoresistance. *Nature Med.* **18**, 1651–1657 (2012).
45. Mignone, J. L., Kukekov, V., Chiang, A. S., Steindler, D. & Enikolopov, G. Neural stem and progenitor cells in nestin-GFP transgenic mice. *J. Comp. Neurol.* **469**, 311–324 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. P. Pandolfi for providing *Pml*^{−/−} mice and D. Rowe, J. Butler and S. Rafii for providing Col2.3-GFP mice. We are grateful to L. Tesfa and O. Uche for technical assistance with sorting. S.P. is supported by a New York Stem Cell Foundation-Druckenmiller Fellowship and C.S. by the German Research Foundation (DFG) (Emmy-Noether-Program). J.A. was supported by Training Grant T32 063754. This work was enabled by the National Institutes of Health (R01 grants DK056638, HL069438, HL097700) to P.S.F.

Author Contributions Y.K. and C.S. performed the whole-mount imaging experiments and analysed the data; I.B. and S.P. performed FACS sorting and qPCR analyses; J.A. and A.B. performed computational modelling and statistical analysis of the data; D.Z. analysed NG2-*cre*^{ERTM}/iDTR mice; T.M. performed the imaging experiments in bone marrow sections; Q.W. and J.C.M. analysed the RNA-seq data; D.L. performed the LTC-IC assay and characterization of endothelial cells by FACS; K.I. provided mice and interpreted data; P.S.F. initiated and directed the study. Y.K. and P.S.F. wrote the manuscript. All of the authors contributed to the design of experiments, discussed the results and commented on the manuscript.

Author Information The RNA sequencing data have been deposited in Gene Expression Omnibus under accession number GSE48764. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.S.F. (paul.frenette@einstein.yu.edu).

METHODS

Animals. B6.Cg-Tg(Cspg4-cre/Esr1*)BAkik/J, STOCK Tg(TIE2GFP) 287Sato/J (TIE2-GFP), C57BL/6-Gt(ROSA)26 Sortm1(HBEGF) Awai/J, B6.Cg-Gt(ROSA)26Sortm14(CAG-tdTomato)Hze/J and B6.129-Leprtm2(cre)Rck/J (Lepr-Cre) mice (all from Jackson Laboratory), C57BL/6-CD45.1/2 congenic strains (from the National Cancer Institute), *Pml*^{-/-} mice (gift from P.P. Pandolfi) and Nes-GFP mice⁴⁵ were bred in our facilities. Col2.3-GFP mice were a gift from D. W. Rowe. All mice were maintained on the C57BL/6J background. Unless indicated otherwise, 7–12-week-old male and female mice were used. No randomization or blinding was used to allocate experimental groups. All experimental procedures were approved by the Animal Care and Use Committees of Albert Einstein College of Medicine.

Immunofluorescence imaging. *In vivo* staining of bone marrow endothelial cells: to specifically stain endothelial cells, fluorescently conjugated antibodies (2 µg per mouse), Alexa Fluor 633 (50 µg per mouse) or Dil-Ac-LDL (1 µg g⁻¹) were intravenously injected. Mice were perfused with PBS, 10 min (for antibodies) or 4 h (for Alexa Fluor 633 and Dil-Ac-LDL) after injection to remove unbound reagents. Sterna and long bones were collected for FACS or immunofluorescence imaging.

Preparations of whole-mount tissues and frozen sections of long bones: femoral or tibial bones were perfusion-fixed and then post-fixed for 30 min in 4% paraformaldehyde (PFA), incubated in 15% and 30% sucrose for cryoprotection and embedded in optical cutting temperature compound (OCT) (Tissue-Tek) at -20 °C. For whole-mount staining, bones were shaved on a cryostat until the bone marrow cavity was fully exposed. Bones were carefully harvested from melting OCT. Frozen sections were prepared according to the Kawamoto method⁴⁶.

Whole-mount tissue preparation of the sternum: sternal bones were collected and transected with a surgical blade into 2–3 fragments. The fragments were bisected sagittally for the bone marrow cavity to be exposed, fixed in 4% PFA.

Immunofluorescence staining of whole-mount tissues and frozen sections of the sterna and long bones: whole-mount tissues and sections were blocked/permeabilized in PBS containing 20% normal goat serum and 0.5% Triton X-100, and stained with primary antibodies for 1–3 days. The tissues were incubated with secondary antibodies for 2 h. Images were acquired using a ZEISS AXIO examiner D1 microscope (Zeiss) with a confocal scanner unit, CSUX1CU (Yokogawa), and reconstructed in three dimensions with Slide Book software (Intelligent Imaging Innovations) or Velocity software (PerkinElmer).

High-speed multichannel fluorescence intravital microscopy (MFIM): we performed MFIM of the calvarial bone marrow as described previously⁴⁷. Bone marrow sinusoids and arterioles were visualized by administration of 10 µg of Rhodamine 6G (R6G) and 50 µg of Alexa Fluor 633, respectively.

Cell sorting and flow cytometry. For sorting of Nes^{per} and Nes^{retic} cells, bone marrow cells from 3–5 mice were pooled for a single experiment. Bone marrow cells were flushed and digested with 1 mg ml⁻¹ collagenase IV (Sigma) and 2 mg ml⁻¹ dispase (Gibco) in Hank's balanced salt solution (Gibco) for 45 min at 37 °C. All cell-sorting experiments were performed using an Aria Cell Sorter (BD Biosciences). Flow cytometric analyses were carried out using an LSRII flow cytometer equipped with FACS Diva 6.1 software (all BD Biosciences). Dead cells and debris were excluded by FCS, SSC and DAPI (4',6-diamino-2-phenylindole) (Sigma) staining profiles. Data were analysed with FlowJo (Tree Star) or FACS Diva 6.1 software.

Antibodies and staining reagents. APC-anti-Gr-1 (RB6-8C5), phycoerythrin (PE)-anti-CD11b (M1/70), biotin-anti-Lineage (TER-119, RB6-8C5, RA3-6B2, M1/70, 145-2C11), fluorescein isothiocyanate (FITC)/PE-anti-CD45.2 (104), PE/PE-Cy7-anti-Ly6A/E (D7), PE/PE-Cy7-anti-CD117 (2B8), biotin-anti-CD48 (HM48-1), biotin-anti-CD41 (MWReg30), FITC/PE/Alexa 647-anti-Ki67 (SolA 15), biotin-anti-VEGFR3 (AFL4) (all from eBioscience), Alexa 647-anti-VE-cadherin (BV13) APC-anti-CD31 (MEC13.3), FITC/PE/PE-Cy7-anti-CD45.1 (A20), PE-anti-CD150 (TC15-12F12.2) (all from Biolegend), PE-anti-Flk1 (Avas 12alpa1) (from BD Biosciences), anti-tyrosine hydroxylase and anti-NG2 (all from Millipore), anti-GFAP (DAKO), rabbit anti-GFP, Alexa Fluor 488/568-anti-rabbit, Alexa Fluor 633 hydrazide and Dil-labelled acetylated low-density lipoprotein (Dil-Ac-LDL) (all from Invitrogen) and Cy3-anti-αSMA antibody, Hoechst 33342, R6G (all from Sigma) were used in this study.

RNA isolation and quantitative PCR. Bone marrow cells were collected from Nes-GFP mice and Nes^{per} and Nes^{retic} cells were sorted according to their specific flow cytometric profiles. Messenger RNA was purified with polyT-conjugated Dynabeads (Invitrogen) and conventional reverse transcription, using the Sprint PowerScript reverse transcriptase (Clontech), was performed in accordance with the manufacturer's protocol. Quantitative PCR was performed with SYBR Green (Roche) on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). A primer concentration of 300 nM was found to be optimal in all cases. The sequences of the oligonucleotides used are included in Supplementary Table 3.

The PCR protocol consisted of one cycle at 95 °C (10 min) followed by 40 cycles of 95 °C (15 s) and 60 °C (1 min). Expression of glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) was used as a control. The average threshold cycle number (Ct) for each tested mRNA was used to quantify the relative expression of each gene; $2^{-Ct(\text{gene}) - Ct(\text{control})}$.

RNA preparation and next-generation sequencing. Total RNA from sorted Nes^{per} and Nes^{retic} cells was extracted using the RNAqueous kit (Life Technologies) and concentrated using RNA Clean and Concentrator columns (Zymo Research Corporation). The integrity and purity of total RNA were assessed using an Agilent Bioanalyzer and OD260/280. Complementary DNA was generated using a Clontech SMARTer Ultra Low RNA kit for Illumina Sequencing (Clontech Laboratories) from 100 pg–1 ng of total RNA. cDNA was fragmented using Covaris (Covaris Inc.), profiled using an Agilent Bioanalyzer, and subjected to Illumina library preparation using NEBNext reagents (New England Biolabs). The quality and quantity and the size distribution of the Illumina libraries were determined by an Agilent Bioanalyzer 2100. The libraries were then submitted for Illumina HiSeq2000 sequencing according to the standard operation procedure. Paired-end 90- or 100-nucleotide reads were generated and checked for data quality using FASTQC (Babraham Institute) and subjected to data analysis using the platform provided by DNAnexus (DNAnexus). For sorting of Nes^{per} and Nes^{retic} cells, bone marrow cells from 5–7 mice were pooled to get a single biological sample. Triplicate biological samples per population from three independent sortings were processed for RNA-seq.

RNA-seq data analysis. Pre-processing of the RNA-seq reads was performed by Orogenetic corp. The RNA-seq reads were aligned using TopHat (version 1.3.3) and levels of gene expression were quantified using Cufflinks (version 1.3.0). Genes that did not pass the quantification status flagged by Cufflinks in either Nes^{retic} or Nes^{per} cell populations were excluded from the analysis. The quantification status represents the success of the Cufflinks algorithm in quantifying expression from the RNA-seq reads that were mapped for a gene. Using Bioconductor, genes were mapped to their Entrez Gene identifiers using the org.Mm.eg.db (version 2.5.0) package and to corresponding pathways based on definitions provided by the Kyoto Encyclopedia of Genes and Genomes (KEGG) via the KEGG.db (version 2.5.0) annotation package. The expression data was transformed by applying a log₂-transformation to the ratio of $(1 + \text{Nes}^{\text{per}} \text{ FPKM values})$ to $(1 + \text{Nes}^{\text{retic}} \text{ FPKM values})$ where the constant value of 1 was added because of the frequency of zero values observed in raw data. Log fold change of Nes^{per}/Nes^{retic} expression was calculated for pathways defined by the KEGG orthology system (http://www.genome.jp/kegg-bin/get_htext?ko00001.keg) and using the HSC niche gene set defined by previous microarray studies^{21,22}. Pathway enrichment of genes showing differential expression between Nes^{per} and Nes^{retic} cells was assessed using a Kolmogorov–Smirnov test. *P* values were adjusted for multiple testing using the Benjamini–Hochberg method which controls the false discovery rate.

EdU treatment and label-retaining assay. *In vivo* EdU (Invitrogen) labelling was performed as previously described with BrdU¹⁰. In brief, Nes-GFP mice were pulsed with 1 mg of EdU for 14 days followed by an EdU-free chase period of 90 days. Long bones were collected and prepared for immunostaining. The frozen bone marrow sections were stained with a Click-iT EdU Imaging Kit (Invitrogen) according to the manufacturer's instructions.

Cell cycle analysis. Cell cycle analysis was performed as described previously¹⁰. In brief, bone marrow cells were stained with surface markers, fixed in 2% PFA in PBS, washed, permeabilized with 0.1% Triton X-100 in PBS, and stained with anti-Ki67 antibody and Hoechst 33342 at 20 µg ml⁻¹ for 30 min. After washing, cells were analysed by LSRII Flow Cytometer (Becton Dickinson).

CFU-F assay. Sorted cells ($1-3 \times 10^2$) were seeded per well in a 12-well adherent tissue culture plate using Phenol Red free α-MEM (Gibco) supplemented with 20% FBS (Hyclone), 10% MesenCult stimulatory supplement (StemCell Technologies) and 0.5% penicillin-streptomycin. Half of the media was replaced after 7 days and at day 14. Cells were stained with Giemsa staining solution (EMD Chemicals).

***In vivo* treatments.** Poly(I:C) (5 mg kg⁻¹) was injected intraperitoneally every other day (two doses). Mice were analysed 24 h after the last injection. 5FU (250 mg kg⁻¹) was injected intravenously, and mice were analysed accordingly. For haematopoietic stem/progenitor cells mobilization, mice received two (day 1), four (day 2) or eight (day 4) consecutive injections of filgrastim (Amgen) (125 µg kg⁻¹ twice a day subcutaneously) or saline.

Induction of Cre^{ERTM}-mediated recombination and iDTR-mediated cell depletion. Male mice older than 8 weeks were injected with 1 mg tamoxifen (Sigma) twice a day for five consecutive days as described previously³⁸. For diphtheria toxin-mediated cell depletion, 100 µg kg⁻¹ of diphtheria toxin (Sigma) were injected intraperitoneally 2 days after the last tamoxifen injection.

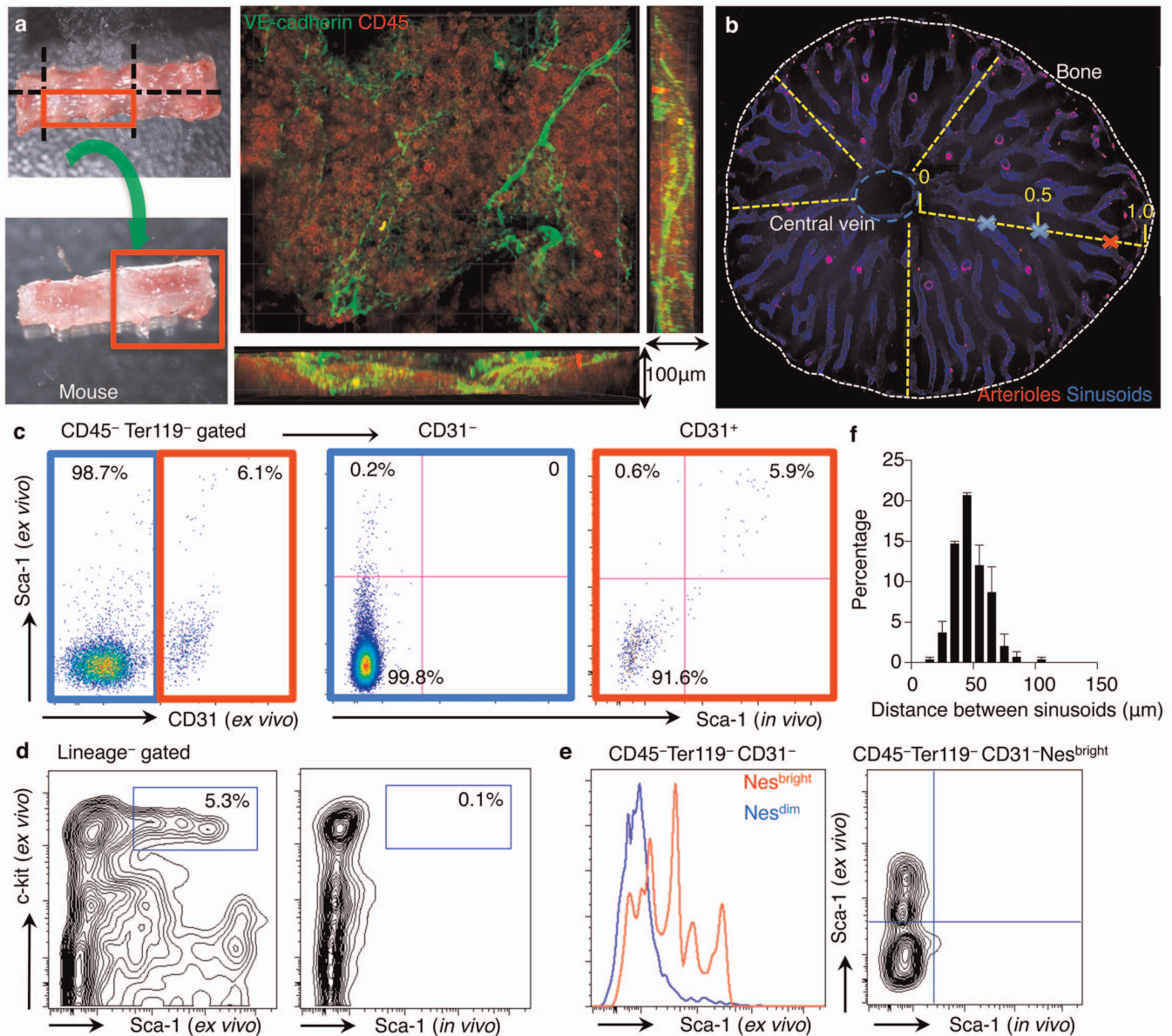
LTC-IC assay. To determine LTC-IC numbers, serial dilutions of bone marrow mononuclear cells were plated on single-well stromal cultures; 4 weeks later each individual well was assayed for the presence of CFU-C as described⁴⁸. Bone marrow HSC frequencies were estimated by the Newton–Raphson method of maximum likelihood and Poisson statistics as the reciprocal of the number of test cells that yielded a 37% negative response. LTC-IC frequencies and statistical significance were determined by using the ELDA software (<http://bioinf.wehi.edu.au/software/elda/>).

Competitive reconstitution. Competitive repopulation assays were performed using the CD45.1/CD45.2 congenic system. Equivalent volumes of bone marrow cells collected from NG2-*cre*^{ERTM}/iDTR or Control mice (CD45.2) were transplanted into lethally irradiated (12 Gy) CD45.1 recipients with 0.3×10^6 competitor CD45.1 cells. CD45.1/CD45.2 chimaerism of recipients' blood was analysed up to 4 months after transplantation.

Computational modelling of HSC localization. Refer to Extended Data Fig. 4. The computational simulation was performed with MATLAB software.

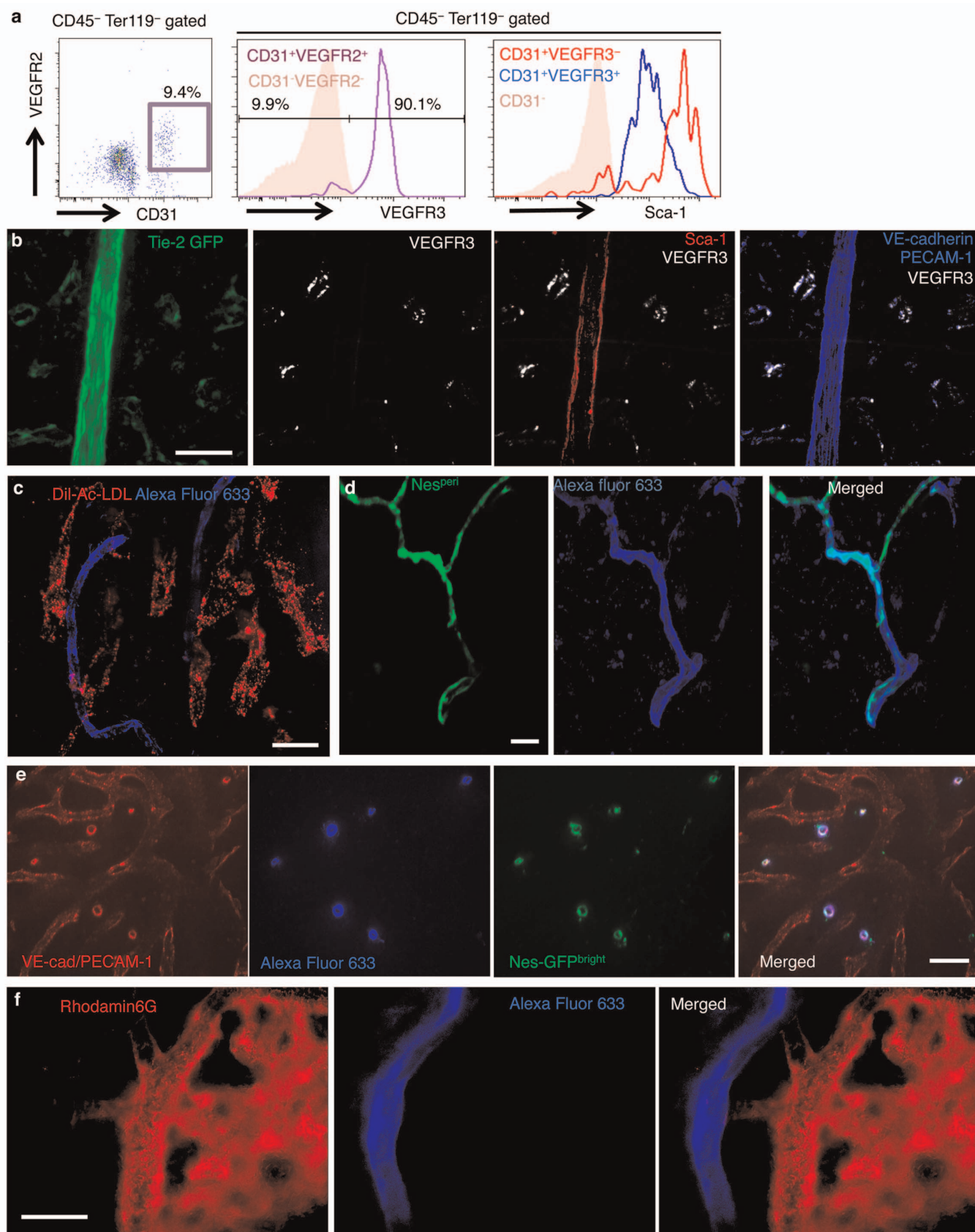
Statistics. All data are represented as mean \pm s.e.m. Comparisons between two samples were done using the unpaired Student's *t* tests. One-way ANOVA analyses followed by Tukey's multiple comparison tests were used for multiple group comparisons. Two-sample Kolmogorov–Smirnov tests were used for comparisons of distribution patterns. Statistical analyses were performed with GraphPad Prism 6 or MATLAB software. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

46. Kawamoto, T. Use of a new adhesive film for the preparation of multi-purpose fresh-frozen sections from hard tissues, whole-animals, insects and plants. *Arch. Histol. Cytol.* **66**, 123–143 (2003).
47. Chiang, E. Y., Hidalgo, A., Chang, J. & Frenette, P. S. Imaging receptor microdomains on leukocyte subsets in live mice. *Nature Methods* **4**, 219–222 (2007).
48. Miller, C. L., Dykstra, B. & Eaves, C. J. Characterization of mouse hematopoietic stem and progenitor cells. *Curr. Protoc. Immunol.* **22**, 22B.2.1–22B.2.31 (2008).



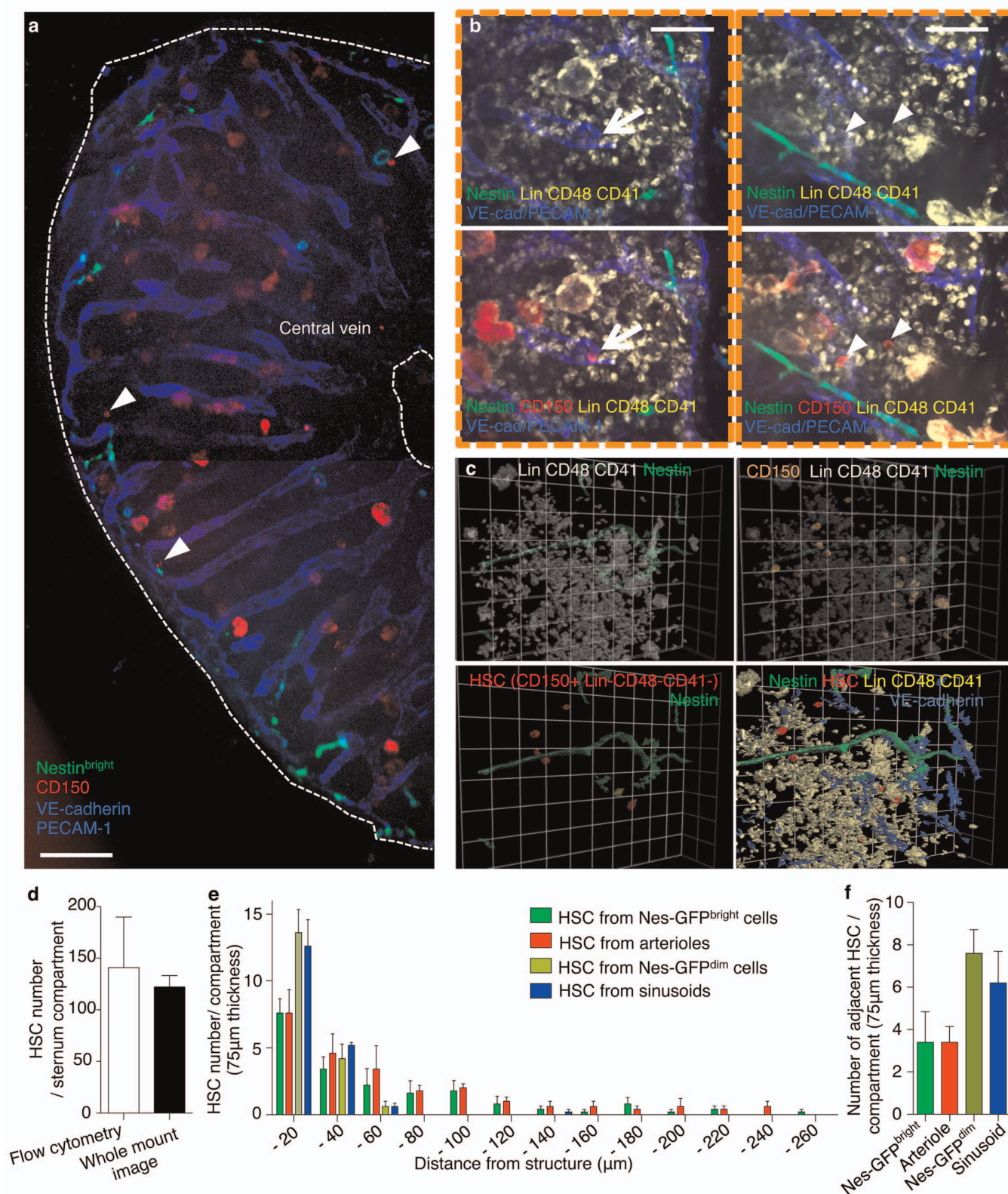
Extended Data Figure 1 | Whole-mount immunofluorescence imaging techniques of the sternal and long-bone bone marrow using *in vivo* immunofluorescence staining of vessels. **a**, Whole-mount immunofluorescence techniques of the sternal bone marrow. Tissue preparation and representative images of the vasculature and hematopoietic cells. **b**, Analysis of the localization of sinusoids, arterioles and HSCs in the femoral bone marrow of transverse-shaved whole-mount immunofluorescence images. The central vein was identified and localization of sinusoids, arterioles

and HSCs were plotted on the axis between central vein and bone as a ratio from 0 to 1. **c**, Representative FACS plots of bone marrow $CD45^- Ter119^-$ stromal cells of three independent experiments. Anti-Sca-1 antibody administered intravenously stains a fraction of $CD31^+$ endothelial cells whereas $CD31^-$ cells are not stained. **d, e**, Haematopoietic (**d**) or Nes-GFP^{bright} mesenchymal (**e**) progenitor cells are not stained by intravenously injected anti-Sca-1 antibody. **f**, Average distances between individual sinusoidal vessels in the femoral bone marrow. $n = 6$ mice.



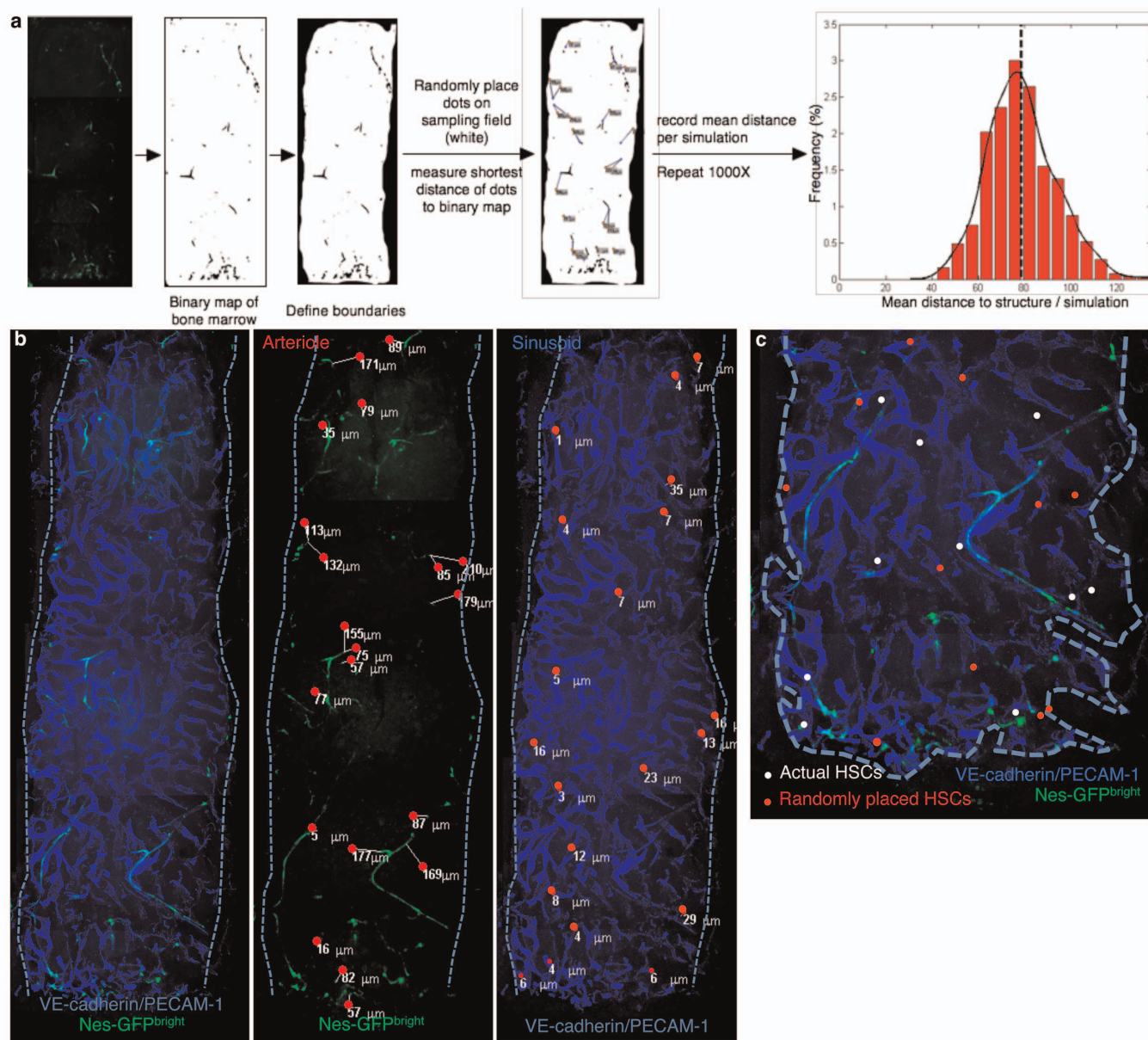
Extended Data Figure 2 | Identification of bone marrow arterioles. **a**, FACS plots of bone marrow endothelial cells. Bone marrow endothelial cells are identified as a VEGFR2⁺ CD31⁺ population. Representative data of 3 mice. ~90% of bone marrow endothelial cells are VEGFR2⁺ VEGFR3⁺ Sca-1^{lo} (sinusoidal) and ~10% are VEGFR2⁺ VEGFR3⁻ Sca-1^{hi} (arteriolar). **b**, Whole-mount images of femoral bone marrow from TIE2-GFP mice stained with anti-VEGFR3, anti-Sca-1, anti-VE-cadherin and anti-PECAM1 antibodies. Scale bar, 25 μ m. **c**, Whole-mount images of the sternal bone marrow stained with Alexa Fluor 633 and Dil-Ac-LDL *in vivo*. Alexa Fluor

633⁺ arteriolar vessels do not take up Dil-Ac-LDL. Scale bar, 50 μ m. **d, e**, Whole-mount images of sternal (**d**) and transverse-shaved femoral (**e**) bone marrow from Nes-GFP mice stained with Alexa Fluor 633 *in vivo* (**d, e**) and anti-PECAM1, anti-VE-cadherin antibodies (**e**). Alexa Fluor 633 specifically stains vessels accompanied by Nes-GFP^{bright} cells (arterioles). Scale bar, 50 μ m. **f**, Intravital imaging of the mouse calvarial bone marrow stained with intravenously injected Rhodamine 6G and Alexa Fluor 633. Sinusoidal vessels identified by Rhodamine 6G are not stained with Alexa Fluor 633. Scale bars, 100 μ m.



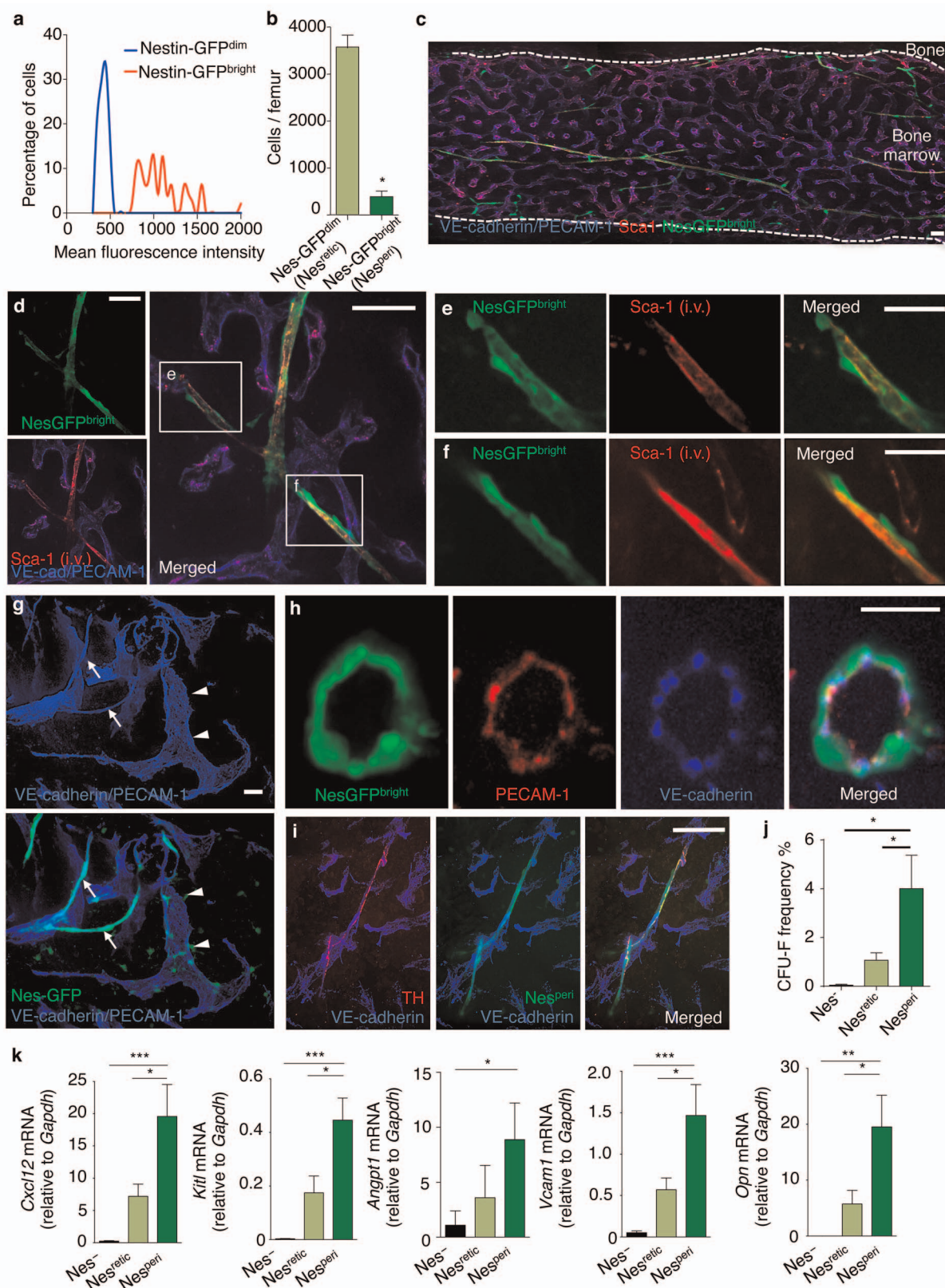
Extended Data Figure 3 | Three-dimensional analysis of sinusoids, arterioles and HSCs by the whole-mount immunofluorescence imaging technique of the bone marrow. **a**, Illustrative example of transverse-shaved femoral bone marrow. Arrowheads denote HSCs. Scale bar, 100 μm. **b, c**, Strategy to identify phenotypic CD150⁺ CD41⁻ CD48⁻ Lineage⁻ HSCs. Megakaryocytes are distinguished by their size and CD41 expression. **b**, Two representative areas highlighted in dashed squares in Fig. 1f are shown in high magnification. Arrowheads denote HSCs, arrows show CD150⁺ Lin⁻ CD48⁻ CD41⁻ cells. Scale bar, 50 μm. **c**, Three-dimensionally reconstructed images.

Grid, 50 μm. **d**, Estimated HSC number per sternal segment measured by FACS and whole-mount image analysis. **e, f**, Distances of HSCs to Nes-GFP^{bright} cells, Nes-GFP^{dim} ($n = 98$ HSCs from 5 mice), arterioles or sinusoids ($n = 119$ HSCs from 5 mice) shown in absolute numbers (**e**) and absolute numbers of adjacent HSCs to those structures (**f**) per sternal segment (75-μm thickness). Similar distribution patterns were obtained when plotting distances of HSCs from Nes-GFP^{peri} cells or arterioles (two-sample Kolmogorov-Smirnov test; $P = 0.97$), and from Nes-GFP^{dim} cells or sinusoids (two-sample Kolmogorov-Smirnov test; $P = 0.45$).



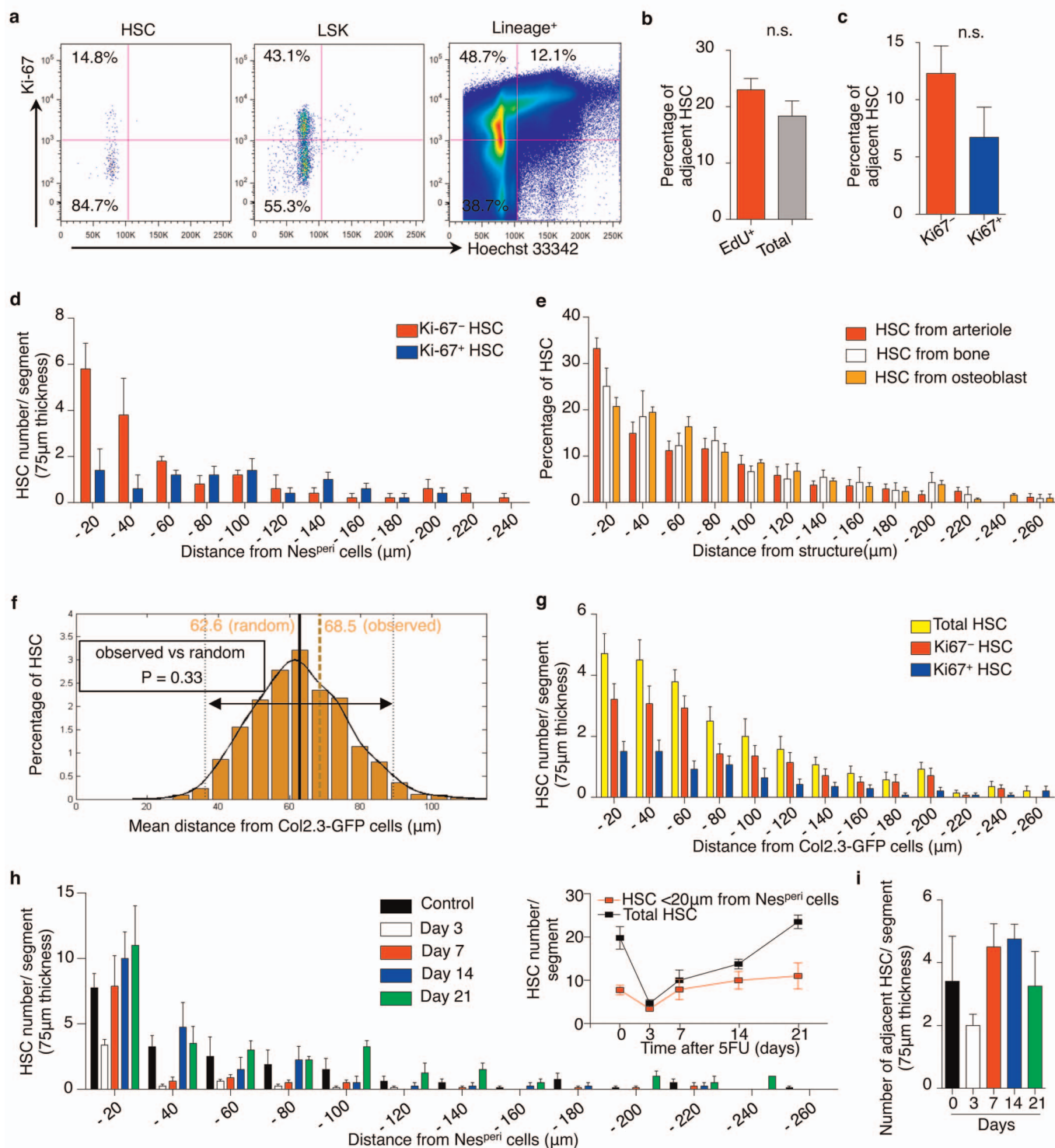
Extended Data Figure 4 | A null modelling of the spatial relationship between HSCs and arteriolar or sinusoidal vessels. **a**, Computational simulation of randomly distributed HSCs on images of whole-mount prepared sterna. To establish the null-model, binary spatial maps of the sinusoids, arterioles (marked by Nes-GFP^{bright} cells, details in Fig. 2a, b and Extended Data Figs 2d, e and 5a–h) were defined from the images of whole-mount prepared sterna. To simulate a null model in which HSCs are not preferentially localized in the marrow, we randomly placed 20 HSCs (to reflect the mean HSCs/sternum observed *in situ*) on the unoccupied regions of the spatial maps and measured the Euclidean distance of HSCs to the nearest vascular structure. The means of 1,000 simulations defined a distribution of mean distances one

would observe for non-preferentially localized HSCs in relation to the respective structures. If the *in situ* distance measurements were not statistically significantly different from those obtained by a random placement of HSCs on the same structures *in silico*, this would indicate a non-preferential spatial HSC distribution. The cumulative probability $P(X \leq \mu)$ of observing the *in situ* mean X was calculated based on the normal distribution, $N(\mu, \sigma^2)$, with mean (μ) \pm s.d. (σ) obtained from our simulation on a map of each bone marrow structure. **b**, **c**, Illustrative examples depicting measurements of randomly placed HSCs (**b**), and distributions of actual (white) and randomly placed (red) HSCs (**c**) on the image in Fig. 1f.



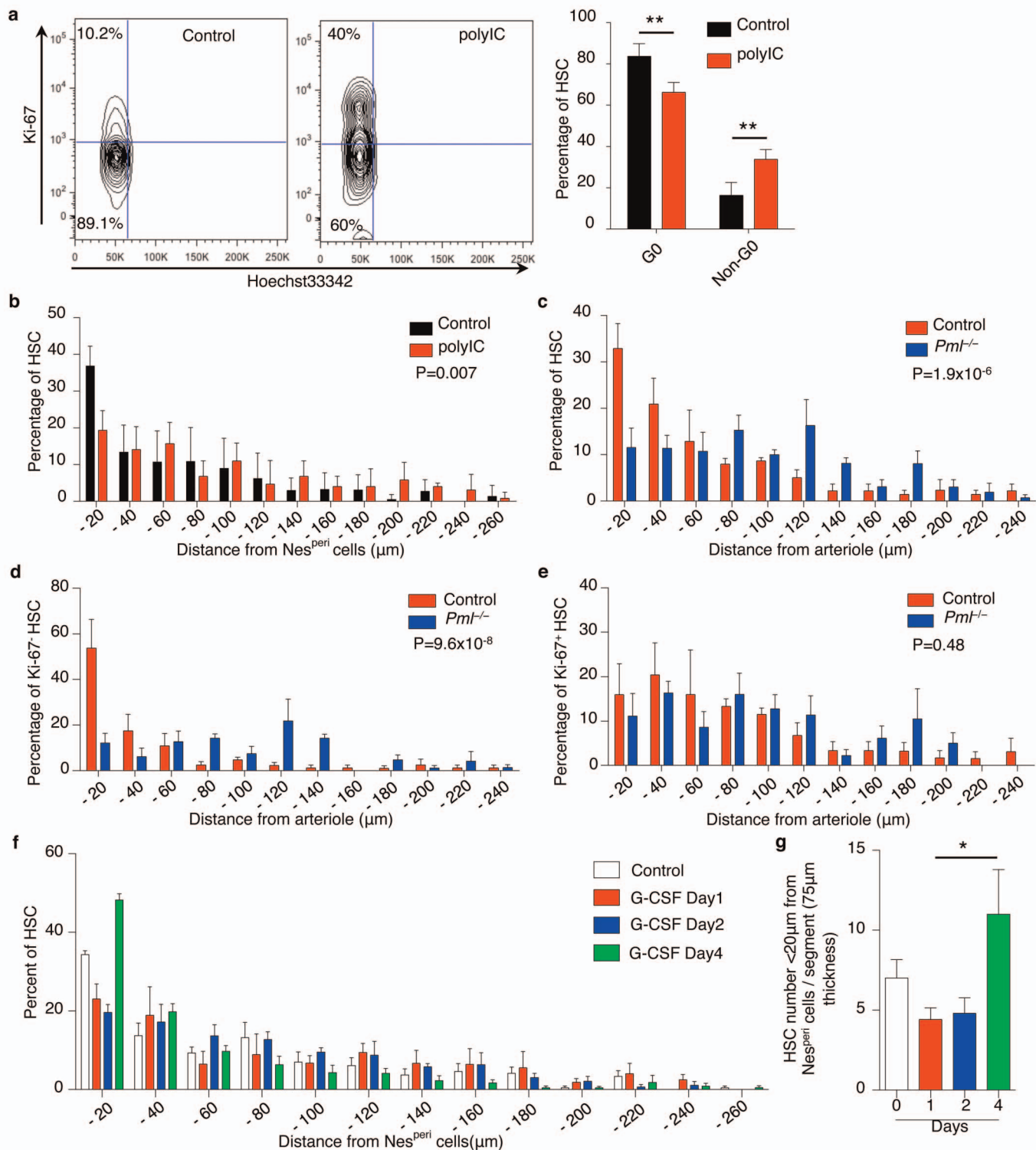
Extended Data Figure 5 | Distinct vascular structures are associated with two types of Nestin⁺ cells. **a**, Quantification of GFP fluorescence intensity of Nes-GFP⁺ cells by whole-mount imaging of the femoral bone marrow. Two distinct populations, Nes-GFP^{bright} and Nes-GFP^{dim}, can be distinguished based on fluorescence intensity. $n = 423$ cells from 4 mice. **b**, The absolute numbers of Nes-GFP^{bright} (Nes^{peri}) and Nes-GFP^{dim} (Nes^{retic}) cells in the femoral bone marrow analysed by FACS. $n = 3$ mice. **c–f**, Whole-mount images of longitudinally shaved femoral bone marrow from Nes-GFP mice stained with anti-PECAM1, anti-VE-cadherin and anti-Sca-1 antibodies *in vivo*. Low-power overview (**c**) and enlarged images (**d–f**). Scale bars, 100 μ m (**c**), 50 μ m (**d**), 25 μ m (**e, f**). **g**, Whole-mount images of sternum from Nes-GFP

mice stained with anti-PECAM1 and anti-VE-cadherin, showing morphological differences of arterioles (arrows) and sinusoids (arrowheads). Scale bar, 25 μ m. **h**, Whole-mount images of transverse-shaved femoral bone marrow from Nes-GFP mice, stained with anti-PECAM1 and anti-VE-cadherin *in vivo*. Scale bar, 10 μ m. **i**, Whole-mount images of the Nes-GFP mouse sternum stained with anti-VE-cadherin and anti-tyrosine hydroxylase (TH) antibodies. Scale bar, 50 μ m. **j**, Quantification of CFU-F content of Nes⁺CD31[−], Nes^{retic} and Nes^{peri} stromal cells. $n = 4$ mice per group. **k**, Quantitative PCR analysis of HSC niche-related genes within Nes^{peri}, Nes^{retic} and Nes-GFP[−]CD31[−] stromal cells. $n = 8, 9, 10, 9, 6$ independent experiments from left to right. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



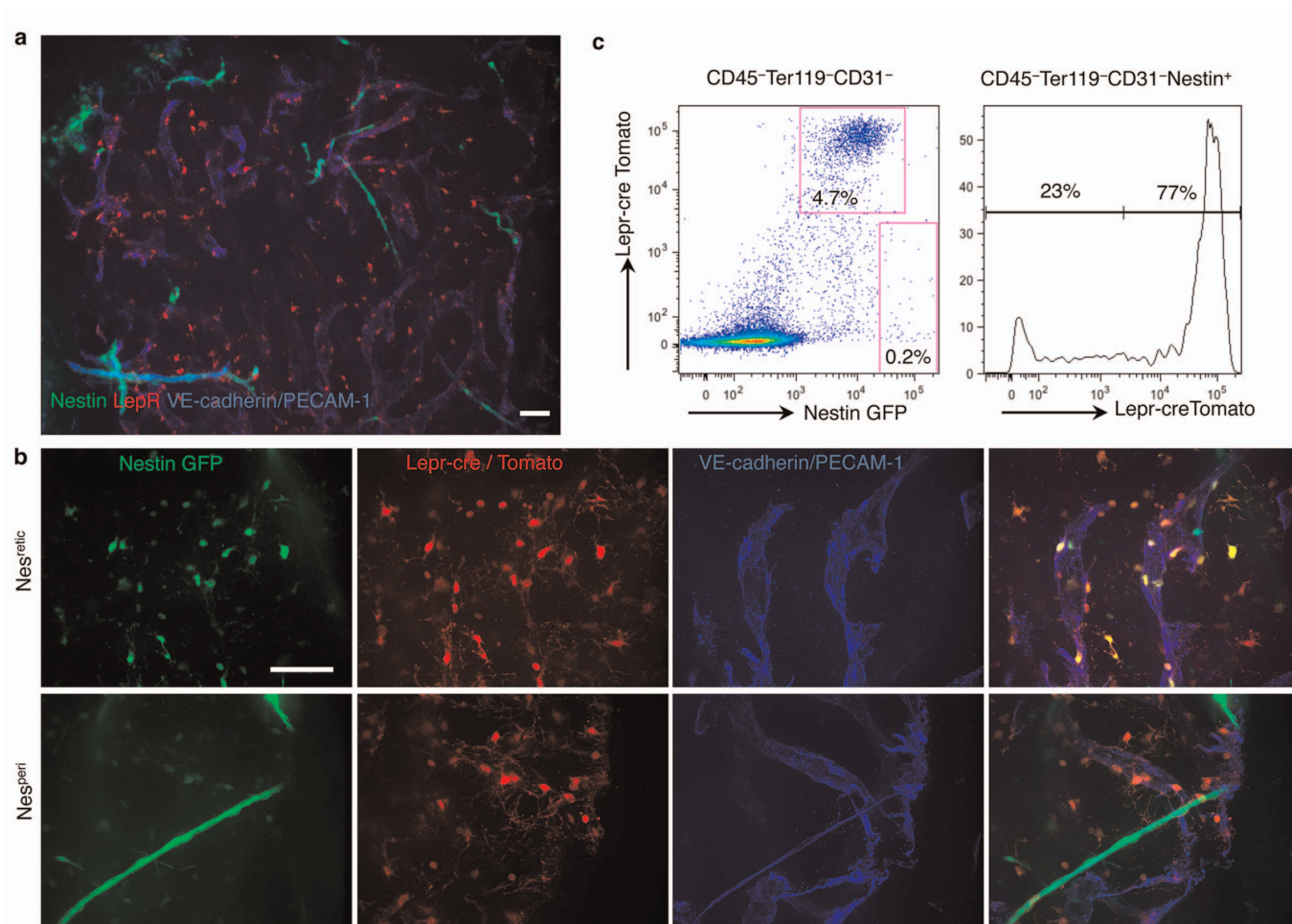
Extended Data Figure 6 | Correlation between cell cycle and localization of HSCs. **a**, FACS plots of HSC cell cycle analysed by staining with anti-Ki-67 antibody and Hoechst 33342. Representative data of 5 mice. **b**, **c**, Percentages of cells adjacent to Nes^{peri} cells in total or EdU⁺ HSCs in the long-bone bone marrow (**b**; $n = 244$, 52 HSCs from 3 mice) and in Ki-67⁻ or Ki-67⁺ HSCs in the sternal bone marrow (**c**; $n = 64$, 116 HSCs from 7 mice). **d**, Localization of Ki-67⁻ and Ki-67⁺ HSCs relative to arterioles in the sternal bone marrow shown as number per sternum segment (75 μ m thickness). $n = 64$, 116 HSCs from 7 mice. **e**, Distributions of distances of HSCs from arterioles, bone or Col2.3-GFP⁺ osteoblasts. $n = 109$, 109, 121 HSCs from 5, 5, 3 mice, respectively. **f**, Probability distribution of the mean distances from 1,000 simulations of 20 randomly positioned HSCs from Col2.3-GFP⁺ cells (Fig. 4e).

A dashed line depicts the actual mean distance observed *in situ*. A solid line and dotted lines show the mean \pm 2 s.d. (95% confidential interval) calculated by the simulation. The observed mean distance of HSCs was not significantly different from that predicted by the simulation ($P = 0.33$). **g**, Localization of total, Ki-67⁻ and Ki-67⁺ HSCs relative to Col2.3-GFP⁺ osteoblasts shown in absolute numbers per sternum segment (75 μ m thickness). $n = 121$, 80, 41 HSCs from 3 mice. **h**, **i**, Localization of HSCs relative to Nes^{peri} cells after 5-FU treatment shown in number per sternum segment (75- μ m thickness). Distribution of HSCs (**h**), kinetics of the number of HSCs located within 20 μ m from Nes^{peri} cells compared to total HSCs (**h**, inset) and adjacent cells (**i**). $n = 98$, 39, 70, 55, 112 HSCs from 5, 9, 7, 4, 4 mice per control, day 3, day 7, day 14 and day 21 groups, respectively.



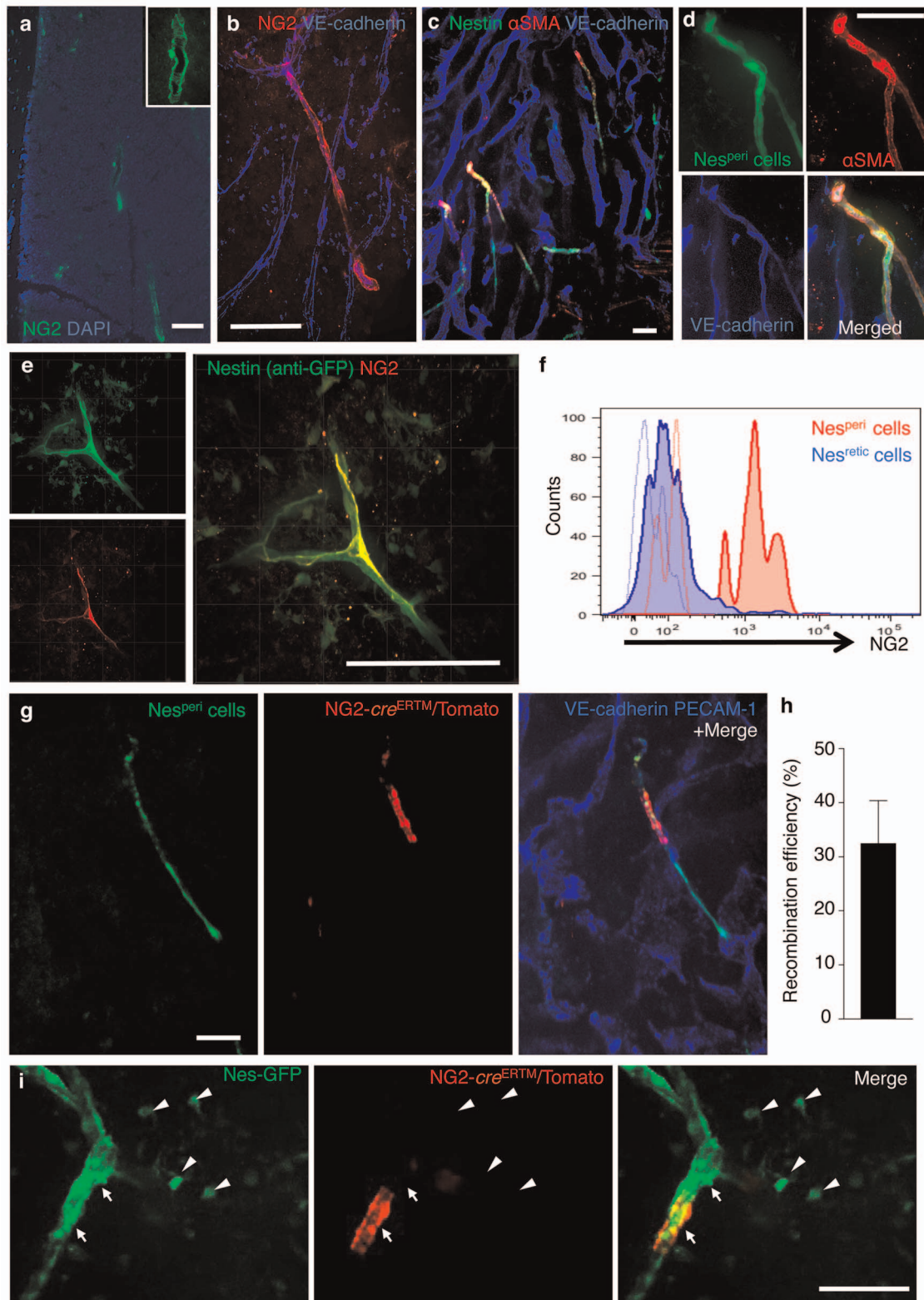
Extended Data Figure 7 | Induction of HSC cell cycle alters their localization. **a**, FACS analysis for HSC (CD150⁺ CD48⁻ Sca-1⁺ c-kit⁺ Lineage⁻ gated) cell cycle by using Ki-67 and Hoechst 33342 staining after Poly(I:C) injection. *n* = 4, 6 mice for control and treatment groups, respectively. **b**, HSC localization relative to Nes^{peri} cells after Poly(I:C) treatment. *n* = 106, 123 HSCs from 9, 4 mice for control and treatment groups, respectively. Two-sample Kolmogorov-Smirnov test; *P* = 0.007. **c**, Altered distances of HSCs from arterioles in *Pml*^{-/-} sternal bone marrow. *n* = 118, 146 HSCs from 5, 4 mice for wild-type control and *Pml*^{-/-} groups, respectively. Wild type/*Pml*^{-/-}: 32.9 ± 5.4%/11.5 ± 4.2% in 0–20-μm proximity, 25.4 ± 3.4%/51.1 ± 6.8% > 80-μm distance. Two-sample Kolmogorov-Smirnov test; *P* = 1.9 × 10⁻⁶. **d**, **e**, Quantification of distances of

Ki-67⁻ quiescent (**d**; *n* = 68, 73 HSCs from 5, 4 mice per wild-type control and *Pml*^{-/-} groups, respectively) and Ki-67⁺ non-quiescent (**e**; *n* = 53, 73 HSCs from 5, 4 mice per wild-type control and *Pml*^{-/-} groups, respectively). HSCs from arteriolar niches in *Pml*^{-/-} bone marrow. Two-sample Kolmogorov-Smirnov test; *P* = 9.6 × 10⁻⁸ (**d**), *P* = 0.48 (**e**). **f**, Localization of HSCs relative to Nes^{peri} cells in the sternal bone marrow after G-CSF treatment. *n* = 98, 73, 122, 94 HSCs from 6, 3, 4, 5 mice for control, day 1, day 2 and day 4 groups, respectively. Two-sample Kolmogorov-Smirnov test; day 1, *P* = 0.026; day 2, *P* = 0.012; day 4, *P* = 0.23 compared to control. **g**, Absolute numbers of HSCs that located within 20-μm distance from Nes^{peri} cells in the sternal bone marrow after G-CSF treatment. *n* = 6, 3, 4, 5 mice for control, day 1, day 2 and day 4 groups, respectively. **P* < 0.05, ***P* < 0.01.



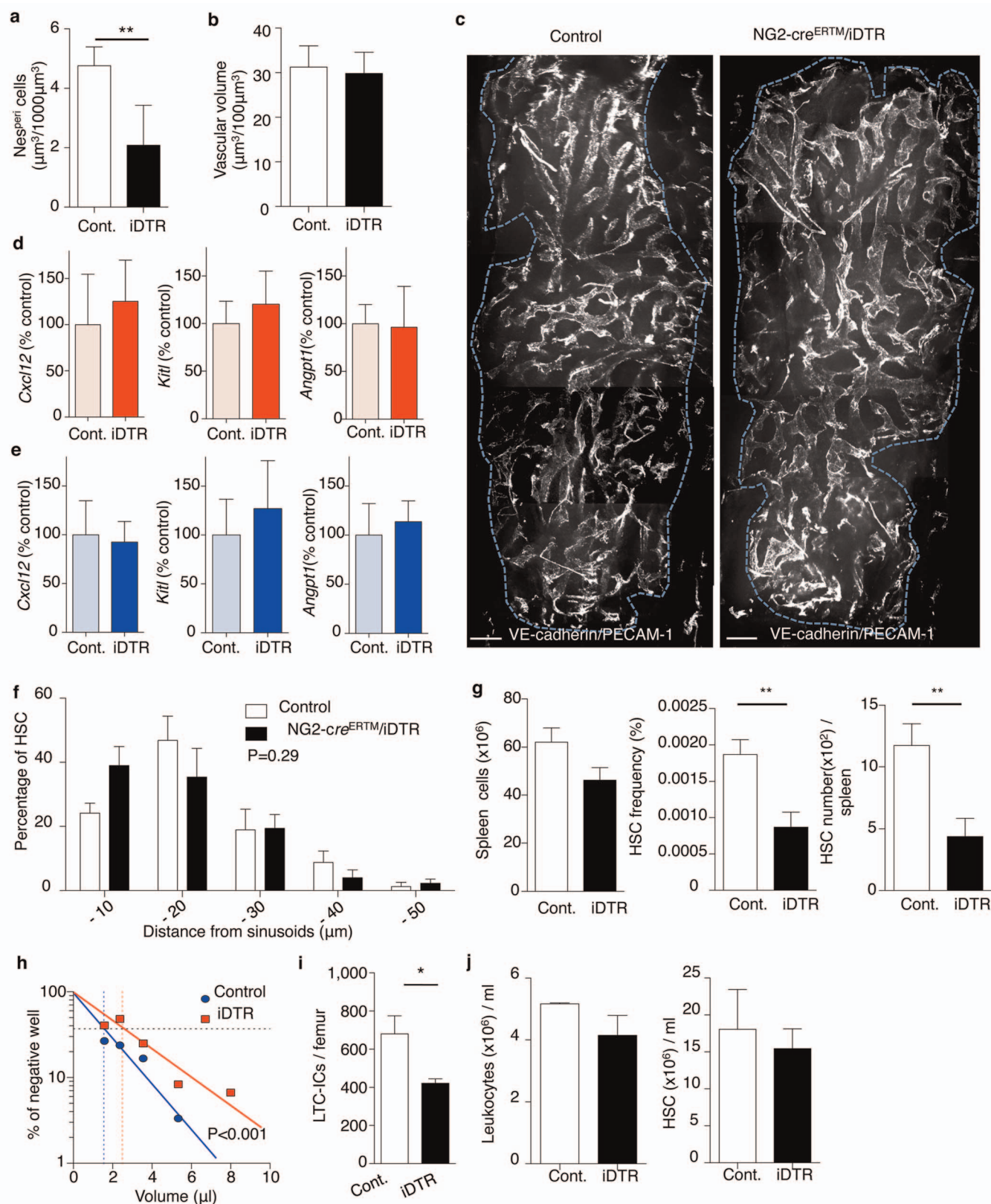
Extended Data Figure 8 | Peri-sinusoidal Nes^{retic} cells, not periarteriolar Nes^{peri}, overlap with LEPR⁺ cells. **a–c**, Analyses of the bone marrow from *Lepr-cre/loxP*-tdTomato/Nes-GFP transgenic mice. Whole-mount

immunofluorescence images of sternum (**a**, **b**) and FACS analysis of the femoral bone marrow (**c**). Representative data of 3 mice. Scale bars, 50 μ m.



Extended Data Figure 9 | Periaarteriolar Nes^{peri} cells are NG2⁺ and α-smooth muscle actin⁺ pericytes. **a, b**, Immunofluorescence images of the bone marrow from wild-type mice stained with anti-NG2 antibody in a sectioned femur (**a**) and a whole-mount sternum (**b**). **c–f**, Whole-mount immunofluorescence (**c–e**) and FACS (**f**) analyses of the bone marrow from Nes-GFP mice stained with anti-α-smooth muscle actin (αSMA) (**c, d**) or

anti-NG2 (**e, f**) antibodies. **g–i**, Analysis of the bone marrow from NG2-cre^{ERTM}/loxP-tdTomato/Nes-GFP transgenic mice. Whole-mount images of sternal bone marrow (**g, i**) and recombination efficiency for tdTomato protein expression on Nes^{peri} cells analysed by whole-mount imaging after tamoxifen treatment (**h**). Arrowheads denote Nes^{retic} cells, arrows show Nes^{peri} cells. $n = 6$ mice. Scale bars, 50 μm.



Extended Data Figure 10 | Periarteriolar NG2⁺ cells are essential for HSC maintenance in the bone marrow and spleen. Functional analyses of NG2⁺ cells using NG2-cre^{ERTM}/iDTR mice. **a–c**, Quantification of Nes^{peri} cells (**a**) and vascular volume (**b**, **c**), accessed by whole-mount imaging of the sternal bone marrow from NG2-cre^{ERTM}/iDTR/Nes-GFP mice. Scale bars, 100 μm . $n = 3$ mice per group. **d**, **e**, *Cxcl12*, *Kitl* and *Angpt1* gene expressions assessed by quantitative PCR in sorted Sca-1^{hi} arteriolar (**d**) and Sca-1^{lo} sinusoidal (**e**) endothelial (CD45⁺ Ter119⁺ CD31⁺) cells after NG2⁺ cell depletion. $n = 4$

mice per group. **f**, HSC localization relative to sinusoids in the sternal bone marrow. $n = 69, 71$ HSCs from 3, 4 mice per control and NG2-cre^{ERTM}/iDTR groups, respectively. Two-sample Kolmogorov–Smirnov test, $P = 0.29$. **g**, Quantification of bone marrow cellularity, frequency and number of phenotypic CD150⁺ CD48⁺ Sca-1⁺ c-kit⁺ Lineage⁺ HSCs in spleen. $n = 6$ mice per group. **h**, **i**, Quantification of long-term reconstituting HSCs by LTC-IC assays. $n = 3$ mice per group. **j**, Numbers of total leukocytes and phenotypic CD150⁺ CD48⁺ Sca-1⁺ c-kit⁺ Lineage⁺ HSCs in blood. $n = 3$ mice per group. * $P < 0.05$, ** $P < 0.01$.

Crystal structure of the 14-subunit RNA polymerase I

Carlos Fernández-Tornero^{1*}, María Moreno-Morcillo^{2*}, Umar J. Rashid², Nicholas M. I. Taylor¹, Federico M. Ruiz¹, Tim Gruene³, Pierre Legrand⁴, Ulrich Steuerwald^{2†} & Christoph W. Müller²

Protein biosynthesis depends on the availability of ribosomes, which in turn relies on ribosomal RNA production. In eukaryotes, this process is carried out by RNA polymerase I (Pol I), a 14-subunit enzyme, the activity of which is a major determinant of cell growth. Here we present the crystal structure of Pol I from *Saccharomyces cerevisiae* at 3.0 Å resolution. The Pol I structure shows a compact core with a wide DNA-binding cleft and a tightly anchored stalk. An extended loop mimics the DNA backbone in the cleft and may be involved in regulating Pol I transcription. Subunit A12.2 extends from the A190 jaw to the active site and inserts a transcription elongation factor TFIIS-like zinc ribbon into the nucleotide triphosphate entry pore, providing insight into the role of A12.2 in RNA cleavage and Pol I insensitivity to α -amanitin. The A49–A34.5 heterodimer embraces subunit A135 through extended arms, thereby contacting and potentially regulating subunit A12.2.

Ribosome biosynthesis is a central cellular process that in eukaryotes requires the coordinated action of three nuclear, DNA-dependent RNA polymerases. Pol I transcribes the ribosomal RNA (rRNA) precursor gene present in several copies in the nucleolus that, in yeast, is subsequently processed into 25S, 18S and 5.8S rRNAs. Pol II, synthesizing messenger RNA, and Pol III, mainly involved in transfer RNA synthesis, contribute to ribosome biogenesis by providing ribosomal protein mRNAs and 5S rRNA, respectively. As Pol I is the most active eukaryotic RNA polymerase contributing up to 60% of the total transcriptional activity¹, rRNA synthesis is a crucial point in cell regulation. Thus, alterations in cell proliferation are accompanied by changes in the synthesis rate of rRNA, and mis-regulation of Pol I has been associated with different types of cancer^{2–4}.

Despite conservation of the overall structure and function, differences between Pol I, Pol II and Pol III probably determine their specific activities⁵. The yeast Pol I enzyme has a total mass of 589 kilodalton (kDa) and consists of 14 subunits⁶, whereas Pol II and Pol III contain 12 and 17 subunits, respectively⁵. All three RNA polymerases share a horseshoe-shaped core built of ten subunits. The Pol I core includes the two largest subunits, A190 and A135, forming the DNA binding cleft, plus five subunits present in all nuclear RNA polymerases (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12) and the AC40–AC19 heterodimer, shared with Pol III and homologous to Rpb3–Rpb11 in Pol II. The core is completed by A12.2, a subunit formed by two zinc ribbons that participates in RNA cleavage⁷. Whereas in Pol III subunit C11 performs a similar function, in Pol II the corresponding amino- and carboxy-terminal zinc ribbons are contributed by subunit Rpb9 and the elongation factor TFIIS, respectively⁸. Outside the core, the A43–A14 heterodimer, related to Rpb7–Rpb4 in Pol II and C25–C17 in Pol III, forms the stalk that provides a platform for initiation factors and interacts with newly synthesized RNA⁷. An additional peripheral heterodimer is formed by subunits A49 and A34.5, the dimerization module of which is related to similar modules in the Pol III C37–C53 heterodimer⁹ and in Pol-II-specific TFIIF¹⁰. Additional homology has been detected between the

tandem winged-helix (tWH) domain in A49 and the Pol III subunit C34 or Pol-II-specific TFIIE^{9,10}.

So far, atomic structures are only available for the yeast Pol II enzyme (reviewed in ref. 11). In contrast, only electron microscopy (EM) reconstructions at about 1-nm resolution have been reported for Pol I (ref. 7) and Pol III (ref. 12). The overall architecture of Pol I was initially revealed by EM studies^{13,14}, whereas crosslinking coupled to mass spectrometry suggested the approximate locations of A12.2 and A49–A34.5 (ref. 15). In addition, crystal structures of the A43–A14 stalk⁷, the A49–A34.5 dimerization module and the A49 tWH domain¹⁰ contributed to a better understanding of Pol I. To shed light on the molecular mechanism of Pol I transcription, we have determined the atomic structure of the complete yeast Pol I enzyme, providing unprecedented insight into its overall architecture and possible functional roles of its components.

Structure determination of yeast Pol I

The crystallographic analysis of yeast Pol I relied on a large-scale purification protocol, initially established for Pol III (refs 12, 16), that yields large amounts of pure, homogeneous and transcriptionally active Pol I. Three related crystal forms varying in cell dimensions were obtained in space group C2 and diffracted beyond 3.5 Å resolution (Extended Data Table 1). Initial molecular replacement phases using the Pol II core¹⁷ as search model allowed the location of heavy atom positions in a Ta₆Br₁₂ derivative that, after SIRAS phasing, yielded a map where the core and stalk could be built. Phasing to higher resolution using a Yb derivative, combined with the partial model, produced a map that allowed the completion of the model (Extended Data Fig. 1a–d). The correct chain tracing was confirmed by a selenomethionine derivative (Extended Data Fig. 1e, f and Extended Data Table 1). Refinement against the best diffracting crystal form, with a resolution of 3.0 Å according to CC_{1/2} > 0.3 criterion¹⁸ (3.27 Å resolution according to the *I*/ σ *I* > 2.0 criterion), yielded a model with good statistics and excellent geometry (Extended Data Table 1). The C-terminal A49 tWH

¹Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040 Madrid, Spain. ²European Molecular Biology Laboratory, Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ³Department of Structural Chemistry, Georg-August-University, Tammannstraße 4, 37077 Göttingen, Germany. ⁴SOLEIL Synchrotron, L'Orme des Merisiers Saint Aubin, 91192 Gif-sur-Yvette, France. [†]Present address: Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany.

*These authors contributed equally to this work.

domain and several other regions are disordered and could not be modelled.

Pol I has a wide cleft and a closed clamp

The crystal structure of Pol I reveals several striking new features compared to other RNA polymerases (Fig. 1a and Extended Data Fig. 2). Although the overall shape of the core is conserved, the Pol I DNA-binding cleft adopts the widest conformation ever reported. In comparison to other RNA polymerases the Pol I cleft is about 10 Å more open (Extended Data Fig. 3). Moreover, superposition of the Pol I structures derived from the three crystal forms allows dividing the enzyme into two modules of similar mass that independently move around five hinges to widen the cleft (Fig. 1b and Extended Data Fig. 3a–c). Interestingly, in all crystal forms the wide Pol I cleft is concomitant with a closed conformation of the clamp, an element that in Pol II has been described to rotate to open and close the DNA-binding cleft¹⁷ (Extended Data Fig. 3b, c). An open clamp has been observed in the Pol II core structure lacking the stalk subunits¹⁷, whereas a closed clamp is present in the complete Pol II enzyme^{19,20} and when bound to

a transcription bubble^{21,22}. In Pol I, the closed conformation of the clamp is probably due to the presence of unique helices in this element that tighten the interaction with subunit Rpb5 (Fig. 1b) and is consistent with the tight association of the Pol I core with the stalk subunits (see below). Accordingly, we speculate that the Pol I clamp always adopts a closed conformation, which might explain the high processivity of this enzyme able to transcribe the long rRNA precursor gene without dissociating from the DNA.

Concurrent with cleft widening, two critical structural elements in Pol I adopt unique conformations. First, the entrance gate to the RNA exit channel is partly occluded by the A190 lid loop (Fig. 1b), whereas the equivalent loop in Pol II and archaeal Pol establishes connections with the wall element above the RNA exit gate^{17,23}. Second, the Pol I bridge helix unwinds at its middle region by a complete helical turn (A190 residues 1012–1016) thereby introducing an additional ~10° kink into the bridge helix compared to other RNA polymerases (Fig. 1b and Extended Data Fig. 3d, e). Partial unfolding of this helix has been proposed to occur during translocation in Pol II (ref. 17) but has only been observed in the bacterial Pol^{24,25}, where in the bacterial

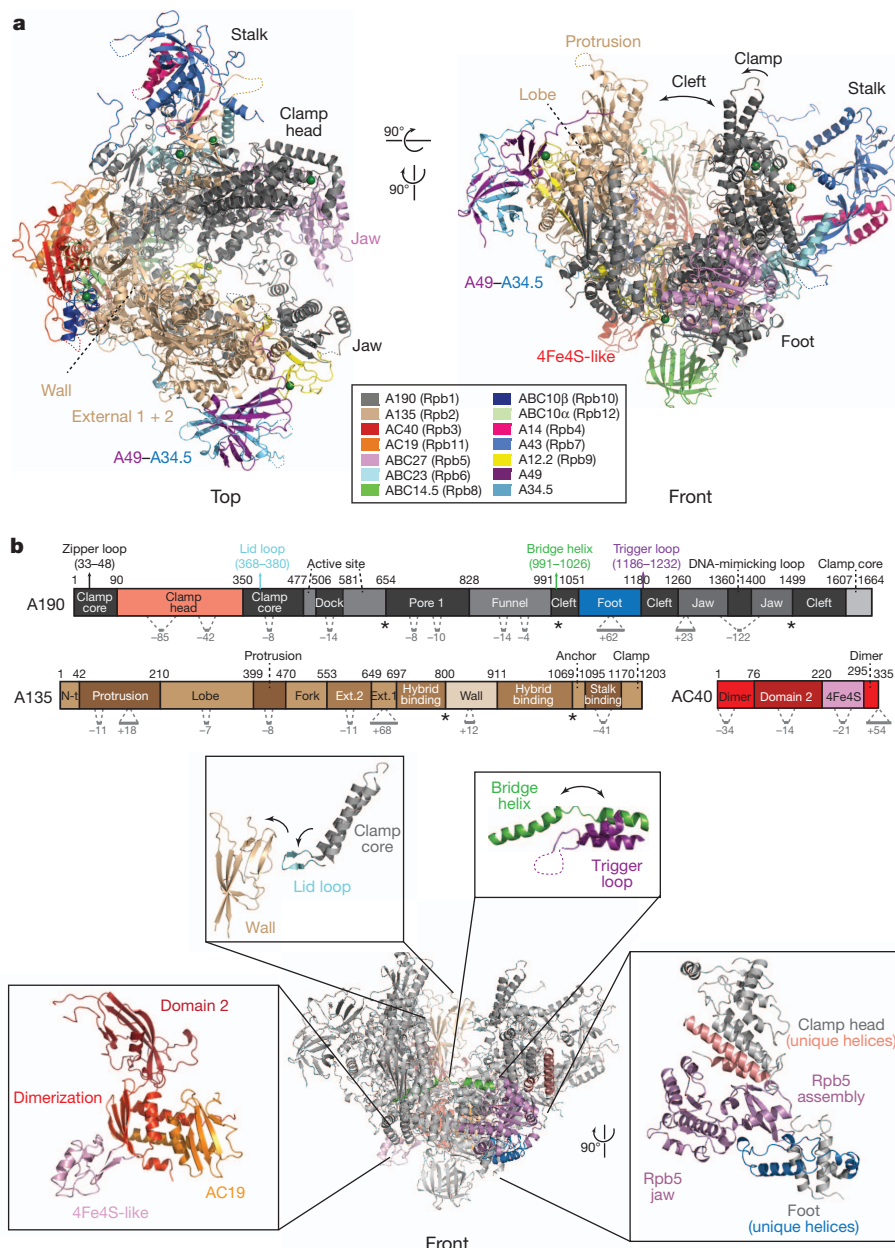


Figure 1 | Crystal structure of yeast RNA polymerase I. **a**, Two orthogonal views of the model with different domains and elements labelled. The colour code for the 14 subunits is given in the box with corresponding Pol II subunits in parentheses. Zn²⁺ ions are shown as green spheres. Arrows indicate the concomitant presence of a wide cleft and a closed clamp. **b**, Close-up views of specific regions in the Pol I core. Colouring of domains according to the bar diagrams. Insertions and deletions in Pol II with respect to Pol I are indicated below the bar diagrams. Black asterisks indicate 'hinges'.

holoenzyme two residues are flipped out of the bridge helix²⁵. Interestingly, these two residues fall into the region that unwinds in Pol I, whereas mutating the corresponding residues in archaeal Pol in some cases increases its transcriptional activity²⁶, presumably because these mutations destabilize the bridge helix (Extended Data Fig. 3e). In accordance with previous systematic mutagenesis studies^{26,27} and molecular dynamics simulations (reviewed in ref. 28), our results also suggest considerable flexibility of the bridge helix that in the case of Pol I even partially unwinds. Different from the bacterial Pol holoenzyme²⁵, the Pol I trigger loop immediately adjacent to the bridge helix is disordered as in the different Pol II structures^{17,21,22} (Extended Data Fig. 3d).

To understand the functional implications of cleft widening better, we have modelled a Pol I elongation complex using the corresponding Pol II structure²². The resulting model (Extended Data Fig. 4a) shows that interactions of subunit A190 with downstream DNA and the beginning of the DNA–RNA duplex are maintained. However, most interactions of subunit A135 with nucleic acids in the vicinity of the active site, critical for NTP selection and translocation, and with the DNA–RNA duplex, can only be established if the Pol I modules move towards each other to close the cleft. This movement presumably involves complete folding of the bridge helix, opening of the RNA exit gate, and the approach of several A135 elements to anchor the transcription bubble (Extended Data Fig. 4b).

Subunits AC40 and AC19 form a heterodimer whose counterparts, Rpb3–Rpb11 in Pol II and archaeal Rpo3–Rpo11, have been described to constitute a platform for the assembly of the enzyme core⁵. AC40 comprises three domains (dimerization, domain 2 and 4Fe4S-like) preceded by a disordered N-terminal tail (Fig. 1b). Although the fold of the first two domains is conserved in other RNA polymerases, AC40 contains a 4Fe4S-like domain that is absent in Pol II but resembles the 4Fe4S cluster domain in Rpo3. However, in contrast to its archaeal homologue, AC40 lacks the four cysteine residues that coordinate the 4Fe4S cluster, further supporting a structural rather than a catalytic role of this domain²³.

An extended loop mimics DNA in the cleft

A notable feature in the Pol I structure is the presence of an extended loop inside the DNA-binding cleft (Fig. 2). We have assigned this loop, which is well ordered in two of the three crystal forms (Extended Data Fig. 5a), to a stretch of ~60 residues in the A190 jaw domain (residues 1340–1400). Although not conserved in Pol II and Pol III, this loop contains several acidic amino acid residues that are conserved within

Pol I enzymes of different species (Extended Data Fig. 5b). Numerous hydrogen bonds and electrostatic interactions are formed between negatively charged residues of the loop and positively charged residues of the cleft. Of note, a patch of acidic residues in this loop (Asp 1385, Glu 1386, Asp 1388, Glu 1389) interacts with Arg 1015 in the unfolded region of the bridge helix. Ordering of the loop also coincides with the stabilization of the C-terminal Zn ribbon of subunit A12.2 (see below) that appears less flexible in the two crystal forms, suggesting that these unique Pol I regions stabilize each other. Although Pol II also contains an insertion and acidic loops in the Rpb1 jaw, they are too short to have a similar role.

The A190 jaw loop contains three short helices and extends from the inner face of the A135 lobe domain to the internal surface of the clamp domain on the opposite side of the cleft. As a consequence, it roughly overlaps with the expected position of the DNA backbone in the transcription bubble when the structure of elongating Pol II²¹ is superimposed with Pol I (Fig. 2). We will therefore refer to it as the ‘DNA-mimicking loop’. Because its position in the cleft is incompatible with elongation, ordering can only occur when the enzyme is not actively transcribing. The fact that this loop is partially disordered in one crystal form and that our Pol I preparation is transcriptionally active (Extended Data Fig. 5c) also suggests that it is a mobile element that only becomes ordered under certain conditions and presumably carries out a regulatory function. For example, it may prevent Pol I from nonspecifically binding nucleic acids or displace the DNA backbone during transcription termination. The DNA-mimicking loop is not essential. A yeast strain lacking most of the loop (residues 1361–1390) is viable but shows a subtle growth defect at 37 °C, but not at 30 °C, compared to the parental strain (see Methods and Extended Data Fig. 5d).

The stalk is tightly anchored to the core

The overall structure of the A43–A14 heterodimer bound to the enzyme core is similar to that determined for the isolated heterodimer⁷. In contrast to Pol II and archaeal Pol, Pol I forms a large interface between the core and the stalk, mainly involving the Pol-I-specific N-terminal tail of A43 (Fig. 3a). This tail forms a hydrophilic helix that provides a strong anchoring point for a Pol-I-specific insertion present at the C-terminal region of A135 (residues 1103–1075) termed stalk-binding domain. This new domain not only wraps around the A43 N-terminal helix in a double-hook fashion, but also contacts residues of the A43 core. In addition, and observed for the first time,

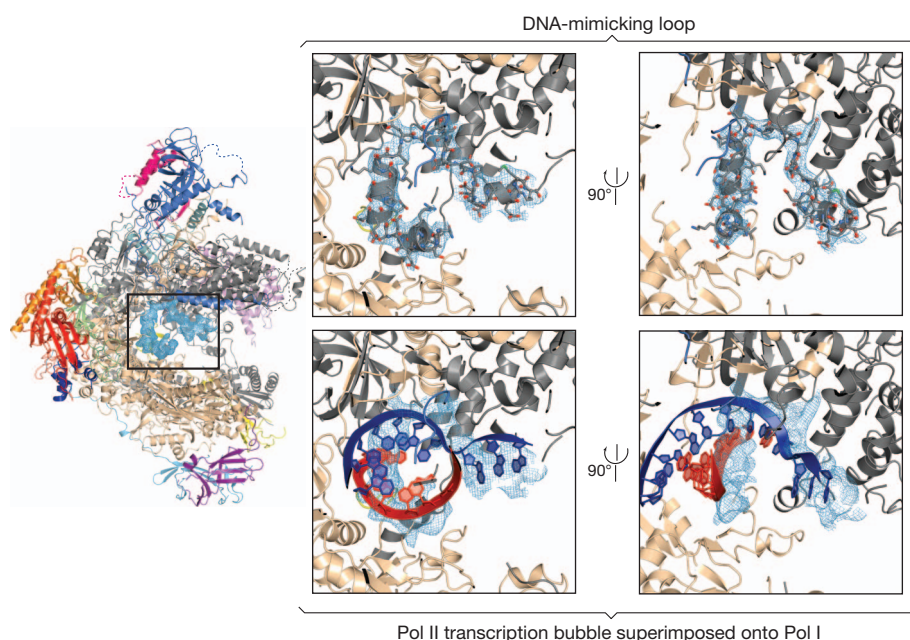


Figure 2 | An extended loop mimics the transcription bubble inside the DNA-binding cleft. Left panel: overall model of Pol I in crystal form C2-90. The σ_A -weighted electron density (contoured at 1σ) corresponds to the extended loop in the DNA-binding cleft. Right panels (top row): two orthogonal orientations of the DNA-mimicking loop in subunit A190 (residues 1361–1399). Right panels (bottom row): superposition of the Pol II elongation complex (Protein Data Bank accession 1I6H) onto the Pol I model. Only the DNA/RNA hybrid is shown with DNA and RNA depicted in blue and red, respectively.

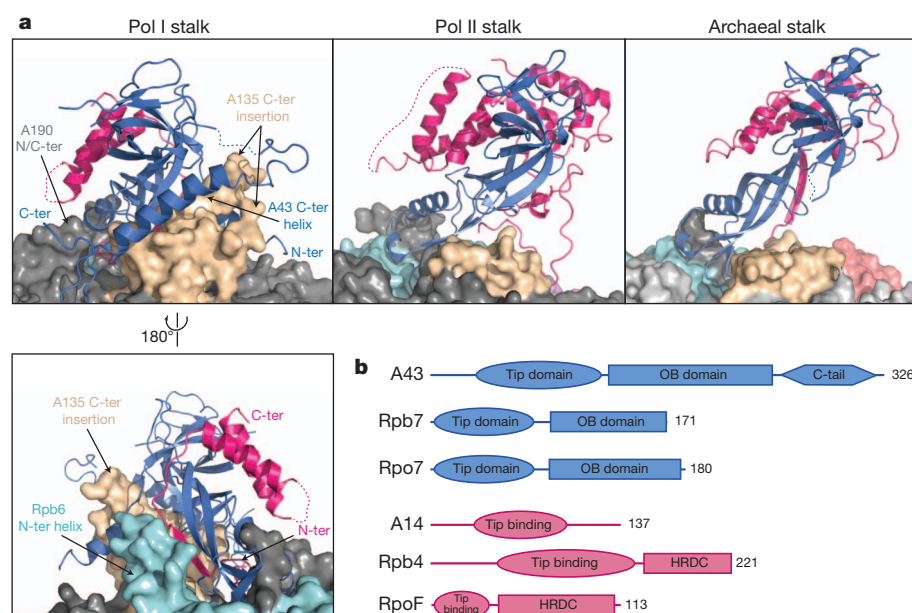


Figure 3 | Stalk anchoring in Pol I, Pol II and archaeal Pol. **a**, Ribbon representations of the Pol I, Pol II and archaeal Pol stalks show differences in their relative orientation with respect to the core (depicted as surface). Main additional connections between the Pol I stalk and the core are labelled. **b**, Schematic representation of the domains found in the stalks of all three polymerases, A43–A14 in Pol I, Rpb7–Rpb4 in Pol II and Rpo7–Rpo4 in archaeal Pol. C-ter, C terminal; HRDC, helicase RNaseD C-terminal domain; N-ter, N terminal; OB, oligonucleotide binding-fold domain.

the N-terminal region of Rpb6 (residues 55–71) adopts a helical conformation that provides additional contacts with A43 and A14. Finally, the A190 N and C termini also contribute to the formation of a well-defined pocket that tightly fixes the stalk. Indeed, in contrast to the dissociable Pol II stalk that can be readily removed from the core under *in vitro* and *in vivo* conditions^{29,30}, the Pol I stalk cannot dissociate even in the presence of moderate urea concentrations³¹. This also argues in favour of a permanently closed conformation of the Pol I clamp, as in Pol II the open clamp conformation has only been observed in the absence of the stalk¹⁷. The tight binding of the Pol I stalk produces a $\sim 10^\circ$ inclination towards the clamp when compared to that of Pol II, whereas the archaeal stalk bends in the opposite direction (Fig. 3). This unique bending may also be critical for the interaction with Pol-I-specific initiation factor Rrn3. The conditional rpa43 mutant (K63E, C118R and Q140R) disrupts the Pol I–Rrn3 interaction³². Whereas residues Lys 63 and Gln 140 are exposed to the solvent and may establish direct contacts with Rrn3 that are compatible with the proposed model of the Pol I–Rrn3 complex³³, residue Cys 118 points to the interior of the tip domain and its mutation rather affects the stalk stability.

In all three crystal forms, the stalk subunit A43 of a neighbouring molecule inserts between the clamp and protrusion domains, establishing contacts with both (Extended Data Fig. 6). The most extensive interaction involves the A43 C-terminal tail (residues 251–326), which folds into a long α -helix followed by a β -hairpin and runs between the A190 lid and zipper loops. These interactions establish a dimer interface around the crystallographic dyad and might contribute to widening the Pol I cleft. Pol I used for this study was monomeric in solution as shown by gel filtration and ultracentrifugation (data not shown). Dimerization of Pol I has been observed by cryo-EM¹³ and in solution under certain conditions³⁴, although its functional relevance is still unclear.

A12.2 reaches the active site

Subunit A12.2 comprises two Zn-binding β -ribbon domains connected by a flexible linker. The N-terminal ribbon is located between the A190 jaw, the A135 lobe and the A49–A34.5 dimerization module, establishing contacts with all three (Fig. 4). Its position is equivalent to that of the Rpb9 N-terminal ribbon in Pol II as previously suggested⁸, but it is slightly shifted and rotated, probably as a consequence of the presence of A49–A34.5. In marked contrast, the A12.2 C-terminal Zn ribbon is located inside the NTP entry pore, as shown in our electron

density maps and validated by zinc anomalous difference maps. This position almost perfectly overlaps with that of the TFIIS C-terminal Zn ribbon in Pol II³⁵ and agrees with experiments showing that A12.2 is required for the intrinsic RNA cleavage activity of Pol I⁷ and the approximate positioning of the C-terminal Zn ribbon by chemical crosslinking¹⁵ (Fig. 4). Indeed, our structure shows that an acidic loop in this subunit lies next to the active site of the enzyme and is probably involved in RNA cleavage, as shown for the corresponding residues in TFIIS³⁶. The two A12.2 Zn ribbons are connected by a linker (residues 45–80) that extends along the bottom face of the enzyme (Extended Data Fig. 7a). The N- and C-terminal portions of this linker follow equivalent paths to those of the Rpb9 and TFIIS linkers, respectively, whereas the middle region of A12.2 (residues 52–70) folds into a helix-containing ‘mini-domain’ that inserts between the A190 funnel and jaw domains. A Pol-I-specific insertion in subunit A190 (residues 863–887) interacts with this A12.2 region through a helix that would clash with the C-terminal domain of Rpb9 in Pol II.

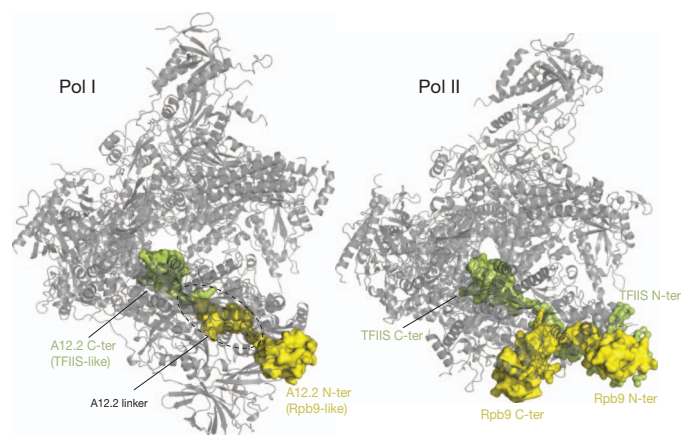


Figure 4 | Pol I subunit A12.2 extends from the jaw to the active site. On the left, A12.2 is depicted in surface representation with the N-terminal Zn ribbon (yellow) occupying the jaw region similar to the N-terminal domain of Rpb9 in Pol II, and its C-terminal part (green) residing in the active site similar to the C-terminal domain of TFIIS. On the right, the Pol II–TFIIS complex (Protein Data Bank accession 1Y1V) with Rpb9 (yellow) and TFIIS (green) depicted in surface representation.

The intrinsic RNA cleavage activity of Pol I has been described to be important for proofreading, termination and rRNA 3' terminal processing during Pol I transcription^{7,37}. Interestingly, a C-terminal Zn ribbon containing a TFIIIS-like acidic loop is also present in subunit C11, the Pol III counterpart for A12.2, indicating a similar domain organization for C11 in Pol III. The position of the C-terminal Zn ribbon in Pol I may also explain the insensitivity of Pol I towards α -amanitin, a strong Pol II inhibitor³⁸. Superposition of the Pol II/ α -amanitin crystal structure with the Pol I structure shows that the inhibitor would need to displace the C-terminal Zn ribbon of A12.2, which is unlikely due to the extensive network of contacts that this domain establishes with the Pol I core (Extended Data Fig. 7b). The medium sensitivity of Pol III towards α -amanitin³⁹ indicates that the C-terminal Zn ribbon of C11 may be less tightly bound to the core Pol III enzyme or may occupy a slightly different position.

Extended arms fix A49–A34.5 on the lobe

The A49–A34.5 heterodimer comprises two folded regions: a dimerization module (A49 residues 6–102 and A34.5 residues 21–141) and the A49 tWH domain (residues 185–415). Whereas the A49 tWH domain seems to be mobile in all three crystal forms and thus cannot be modelled, the dimerization module forms a triple β -barrel—like its *Candida albicans* orthologue¹⁰—that docks on the lobe and external two domains of subunit A135, and also contacts the N-terminal Zn ribbon of subunit A12.2 in an unforeseen manner (Fig. 5). Its position, unambiguously confirmed by selenomethionine anomalous difference maps (Extended Data Fig. 1e), is roughly coherent with that observed in early cryo-EM studies⁴⁰ and by crosslinking experiments¹⁵. Moreover, on binding to the Pol I core, previously unobserved arms protrude from the C-terminal ends of both A49 and A34.5 dimerization domains. Notably, the A34.5 C-terminal tail has an essential role in heterodimer anchoring, as it extends by more than 70 Å along the outer face of A135 and reaches into a cavity at the junction of subunits A135, Rpb10 and AC40 (Fig. 5). Additionally, the first 11 residues in the A49 linker (residues 103–113) strongly interact with the A135 lobe domain and the A12.2 N-terminal Zn ribbon, further anchoring the heterodimer to the core (Fig. 5). These results are coherent with biochemical data showing that either the A49 linker or the A34.5

C-terminal tail is necessary for stable binding of the heterodimer to the core¹⁰. Moreover, interaction with A12.2 could explain the stimulatory role of A49–A34.5 in the intrinsic RNA cleavage of Pol I¹⁰, as stabilization of the A12.2 N-terminal ribbon may also determine the position of its C-terminal ribbon in the cleft.

In Pol III, the C37–C53 heterodimer^{9,10,12} seems to be the equivalent to the A49–A34.5 heterodimer, and the Pol I crystal structure can be perfectly placed into the cryo-EM reconstruction of Pol III¹² (Extended Data Fig. 8a). Nevertheless, as the domain organization is different in these subunits, the N- and C-terminal extensions of C37–C53 probably interact differently with the Pol III core. Our results can also be extended to the Pol II system, where an equivalent dimerization module is present in the TFIIF Rap74–Rap30 heterodimer (Extended Data Fig. 8b, c). Crosslinking^{41,42} and cryo-EM studies⁴³ of Pol II in complex with TFIIF show a similar location for this module and the cryo-EM studies also locate the C-terminal winged helix (WH) domain of Rap30, the A34.5 counterpart, between the protrusion and lobe domains of Pol II. In Pol I, these domains are closer together and interact directly, thereby occluding the corresponding space, consistent with the absence of a WH domain in A34.5. Finally, the Pol I A49 tWH domain probably corresponds to the small subunit of the TFIIE heterodimer, which has also been observed in cryo-EM studies⁴³. Whereas the A49 tWH domain is disordered in the Pol I crystal structure, we speculate that in a pre-initiation complex it may occupy an equivalent position over the transcription start site, in accordance to crosslinking data¹⁵, thus participating in DNA melting as well as in the recruitment of the A43-interacting transcription factor Rrn3 (ref. 44).

The crystal structure of Pol I further corroborates the emerging view that the basic architecture of Pol I, Pol II and Pol III and their specific pre-initiation complexes is conserved⁵. The Pol I crystal structure presented here provides a structural framework to rationalize available genetic and biochemical data on Pol I transcription of ribosomal RNA genes and represents an important step towards a better understanding of eukaryotic transcription mechanisms. In addition, it will hopefully stimulate future research to target the Pol I transcription machinery by small molecule effectors as potential anticancer drugs.

METHODS SUMMARY

Endogenous *Saccharomyces cerevisiae* Pol I containing all 14 subunits was purified using several chromatographic steps. Crystallization using the sitting-drop method yielded crystals diffracting up to 3.0 Å resolution, according to recently defined criteria¹⁸. The Pol I crystal structure was solved by combining phase information from a molecular replacement solution using Pol II as search model with phases obtained from various heavy-atom derivatives, followed by solvent flattening and multi-crystal averaging. The resulting model was refined at 3.03 Å resolution with $R_{\text{work}}/R_{\text{free}} = 19.7/22.9\%$ and excellent geometry.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 May; accepted 4 September 2013.

Published online 23 October 2013.

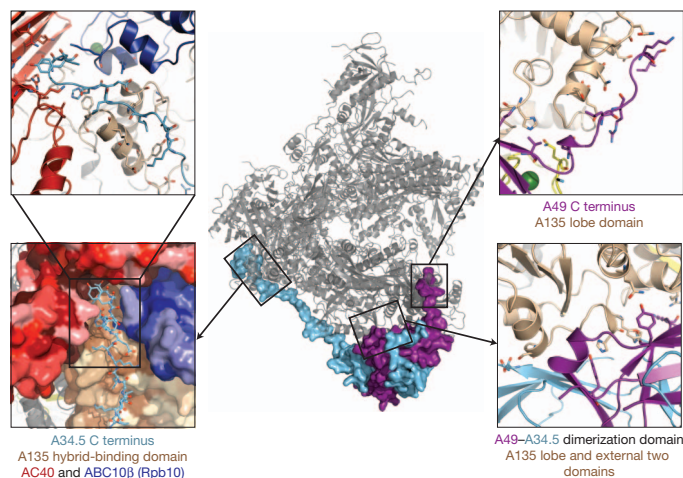


Figure 5 | Anchoring of the A49–A34.5 heterodimer onto the Pol I core. Central panel: A49 (purple) and A34.5 (cyan) are highlighted by surface representation. Left and right panels show residues of the A49–A34.5 heterodimer interacting with the core. Left panels: the C-terminal A34.5 hook inserts into a conserved cavity formed by subunits A135, AC40 and Rpb10. Subunits forming the cavity are coloured according to sequence conservation with darkest colour corresponding to the most conserved residues. Right panel, top: A49 linker residues interact with the lobe domain of A135. Right panel, bottom: A49–A34.5 dimerization domain contacts lobe and external two domains of A135.

- Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
- Drygin, D., Rice, W. G. & Grummt, I. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu. Rev. Pharmacol. Toxicol.* **50**, 131–156 (2010).
- Grummt, I. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev.* **17**, 1691–1702 (2003).
- Moss, T., Langlois, F., Gagnon-Kugler, T. & Stefanovsky, V. A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell. Mol. Life Sci.* **64**, 29–49 (2007).
- Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
- Russell, J. & Zomerdijs, J. C. The RNA polymerase I transcription machinery. *Biochem. Soc. Symp.* **73**, 203–216 (2006).
- Kuhn, C. D. *et al.* Functional architecture of RNA polymerase I. *Cell* **131**, 1260–1272 (2007).

8. Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D. & Cramer, P. Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J. Biol. Chem.* **286**, 18701–18707 (2011).
9. Carter, R. & Drouin, G. The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol. Biol. Evol.* **27**, 1035–1043 (2010).
10. Geiger, S. R. *et al.* RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol. Cell* **39**, 583–594 (2010).
11. Cramer, P. *et al.* Structure of eukaryotic RNA polymerases. *Ann. Rev. Biophys.* **37**, 337–352 (2008).
12. Fernández-Tornero, C. *et al.* Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J.* **29**, 3762–3772 (2010).
13. Bischler, N. *et al.* Localization of the yeast RNA polymerase I-specific subunits. *EMBO J.* **21**, 4136–4144 (2002).
14. Schultz, P., Celia, H., Riva, M., Sentenac, A. & Oudet, P. Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. *EMBO J.* **12**, 2601–2607 (1993).
15. Jennebach, S., Herzog, F., Aebersold, R. & Cramer, P. Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res.* **40**, 5591–5601 (2012).
16. Fernández-Tornero, C. *et al.* Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol. Cell* **25**, 813–823 (2007).
17. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876 (2001).
18. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
19. Armache, K. J., Kettenberger, H. & Cramer, P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl Acad. Sci. USA* **100**, 6964–6968 (2003).
20. Bushnell, D. A. & Kornberg, R. D. Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl Acad. Sci. USA* **100**, 6969–6973 (2003).
21. Gnat, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
22. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004).
23. Hirata, A., Klein, B. J. & Murakami, K. S. The X-ray crystal structure of RNA polymerase from *Archaea*. *Nature* **451**, 851–854 (2008).
24. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
25. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719 (2002).
26. Tan, L., Wiesler, S., Trzaska, D., Carney, H. C. & Weinzierl, R. O. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J. Biol.* **7**, 40 (2008).
27. Jovanovic, M. *et al.* Activity map of the *Escherichia coli* RNA polymerase bridge helix. *J. Biol. Chem.* **286**, 14469–14479 (2011).
28. Weinzierl, R. O. The bridge helix of RNA polymerase acts as a central nanomechanical switchboard for coordinating catalysis and substrate movement. *Archaea* **2011**, 608385 (2011).
29. Edwards, A. M., Kane, C. M., Young, R. A. & Kornberg, R. D. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J. Biol. Chem.* **266**, 71–75 (1991).
30. Mosley, A. L. *et al.* Quantitative proteomics demonstrates that the RNA polymerase II subunits Rpb4 and Rpb7 dissociate during transcription elongation. *Mol. Cell. Proteomics* **12**, 1230–1238 (2013).
31. Huet, J., Buhler, J. M., Sentenac, A. & Fromageot, P. Dissociation of two polypeptide chains from yeast RNA polymerase A. *Proc. Natl Acad. Sci. USA* **72**, 3034–3038 (1975).
32. Peyroche, G. *et al.* The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J.* **19**, 5473–5482 (2000).
33. Blattner, C. *et al.* Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev.* **25**, 2093–2105 (2011).
34. Milkereit, P., Schultz, P. & Tschochner, H. Resolution of RNA polymerase I into dimers and monomers and their function in transcription. *Biol. Chem.* **378**, 1433–1443 (1997).
35. Kettenberger, H., Armache, K. J. & Cramer, P. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).
36. Jeon, C., Yoon, H. & Agarwal, K. The transcription factor TFIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proc. Natl Acad. Sci. USA* **91**, 9106–9110 (1994).
37. Prescott, E. M. *et al.* Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc. Natl Acad. Sci. USA* **101**, 6068–6073 (2004).
38. Keding, C., Gniazdowski, M., Mandel, J. L. Jr, Gissinger, F. & Chambon, P. α -Amanitin: a specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus. *Biochem. Biophys. Res. Commun.* **38**, 165–171 (1970).
39. Weinmann, R. & Roeder, R. G. Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. *Proc. Natl Acad. Sci. USA* **71**, 1790–1794 (1974).
40. De Carlo, S., Carles, C., Riva, M. & Schultz, P. Cryo-negative staining reveals conformational flexibility within yeast RNA polymerase I. *J. Mol. Biol.* **329**, 891–902 (2003).
41. Chen, Z. A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
42. Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
43. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
44. Beckouet, F. *et al.* Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol. Cell. Biol.* **28**, 1596–1605 (2008).

Acknowledgements We thank H. Grötsch for preparing the loop deletion yeast strain, and G. von Scheven and A. Scholz for technical assistance. We are also grateful to C. Vonrhein, G. Bricogne, S. Glatt and A. Romero for advice and discussions. We thank staff from the European synchrotrons SOLEIL, DESY, ESRF and SLS, where data were collected during different stages of the project. In particular, we thank A. Thompson for access and support at beamline Proxima 1 (SOLEIL) and T. Schneider and G. Bourenkov at beamline P14 (PETRA III). We also acknowledge support by the EMBL Heidelberg Protein Expression and Purification, Proteomics Core Facilities and Crystallization Platform, and the ‘Fermentation et culture de microorganismes’ (IFR88, CNRS). We are grateful to M. Bauzan, E. Poilpre and J. Scheurich for yeast fermentation. M.M.-M. and U.J.R. were supported by EMBO Long-Term fellowships, M.M.-M. by the Marie-Curie fellowship (FP7-PEOPLE-2011-IEF 301002), N.M.I.T. by a Fundación Futuro fellowship, F.M.R. by an ESF/CSIC funded JAE-DOC contract and T.G. by the Volkswagen Stiftung via the Niedersachsenprofessur of Prof. G. M. Sheldrick. This work was also partly funded by grant BFU2010-16336 of the Spanish Ministry of Science.

Author Contributions C.F.-T. and C.W.M. initiated the project. C.F.-T. and U.S. established the Pol I purification and obtained Pol I crystals. C.F.-T., M.M.-M. and U.J.R. further improved the Pol I crystals, and collected data and obtained experimental phase information with the help of N.M.I.T., T.G. and P.L. C.F.-T., M.M.-M., N.M.I.T., F.M.R., T.G. and P.L. carried out the crystallographic analysis and model refinement. C.F.-T., M.M.-M. and C.W.M. wrote the manuscript with input from the other authors.

Author Information Atomic coordinates and structure factors of the three related Pol I crystal forms have been deposited at the Protein Data Bank under accession numbers 4C3H, 4C3I and 4C3J. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.F.-T. (cftornero@cib.csic.es) or C.W.M. (cmueller@embl.de).

METHODS

Yeast Pol I expression and purification. The Pol I enzyme was isolated from *Saccharomyces cerevisiae* strain SC1613 (also called YPR110c, provided by Cellzome AG), modified to express endogenous AC40 fused with a C-terminal TAP-tag. Yeast cells were grown overnight in YPDA medium at 30 °C and 180 r.p.m. under controlled conditions and collected at an OD₆₀₀ of 5–6. The cell paste was resuspended in a buffer containing 250 mM Tris-HCl, pH 8, 40% glycerol, 250 mM (NH₄)₂SO₄, 1 mM EDTA, 10 mM MgCl₂, 10 µM ZnCl₂, 12 mM β-mercaptoethanol in the presence of a protease inhibitor cocktail (complete EDTA-free, Roche) and lysed with glass beads in a BeadBeater (BioSpec). After centrifugation at 14,000 r.p.m. for 1 h at 4 °C, the protein lysate was loaded on heparin-sepharose resin (GE Healthcare). The complex was eluted from the resin using high-salt buffer with 1 M (NH₄)₂SO₄ and further incubated with IgG Sepharose (GE Healthcare) for 6 h. After washing, IgG beads were mixed with tobacco etch virus (TEV) protease and incubated overnight at 4 °C. TAP-tag cleaved Pol I was recovered and subsequently purified by ionic exchange on a Mono-Q column (GE Healthcare). The pure Pol I enzyme was concentrated to 7 mg ml⁻¹ in 15 mM Tris-HCl, pH 8.0, 150 mM NaCl and 10 mM dithiothreitol (DTT). Yeast cells containing selenomethionine-labelled Pol I were prepared from an initial glycerol stock pre-adapted to 80 µg ml⁻¹ of selenomethionine. A synthetic complete medium, which included yeast nitrogen base lacking amino acids, D-glucose and an amino acid mix containing different concentrations of selenomethionine, was used for fermentation at 30 °C and 180 r.p.m. Purification was performed under the same conditions as for the native complex.

Yeast Pol I crystallization. Well-diffracting plate-shaped crystals grew during 4–7 days in two different conditions to a maximum size of 0.9 × 0.4 × 0.3 mm³ at 18 °C using vapour diffusion, sitting-drop plates. Protein and crystallization solutions were mixed in a ratio 1:1. In the first condition (16–30% ethylenglycol (EG) and 100 mM 2-(N-morpholino)-ethane-sulphonic acid (MES) pH 6.3–6.9) cryo-protection was reached by stepwise increases of the concentration of EG up to 30% in the reservoir. The second condition included 1–12.5% 2-methyl-2,4-pentanediol (MPD) and 100 mM Tris-HCl, pH 6.3 to 6.9. In this case, crystals were directly soaked in a cryo-solution containing 22.5% MPD. Crystals were frozen in cryo-loops under a nitrogen stream at 100 K. For initial, low-resolution phasing, crystals grown in the presence of EG were soaked for 1 h in a solution containing 2 mM Ta₆Br₁₂. For optimized experimental phasing, an ytterbium derivative crystal was obtained by soaking in a 100 mM Yb-HPDO3A containing solution for 2 min. Selenomethionine-labelled Pol I crystals were grown in similar conditions.

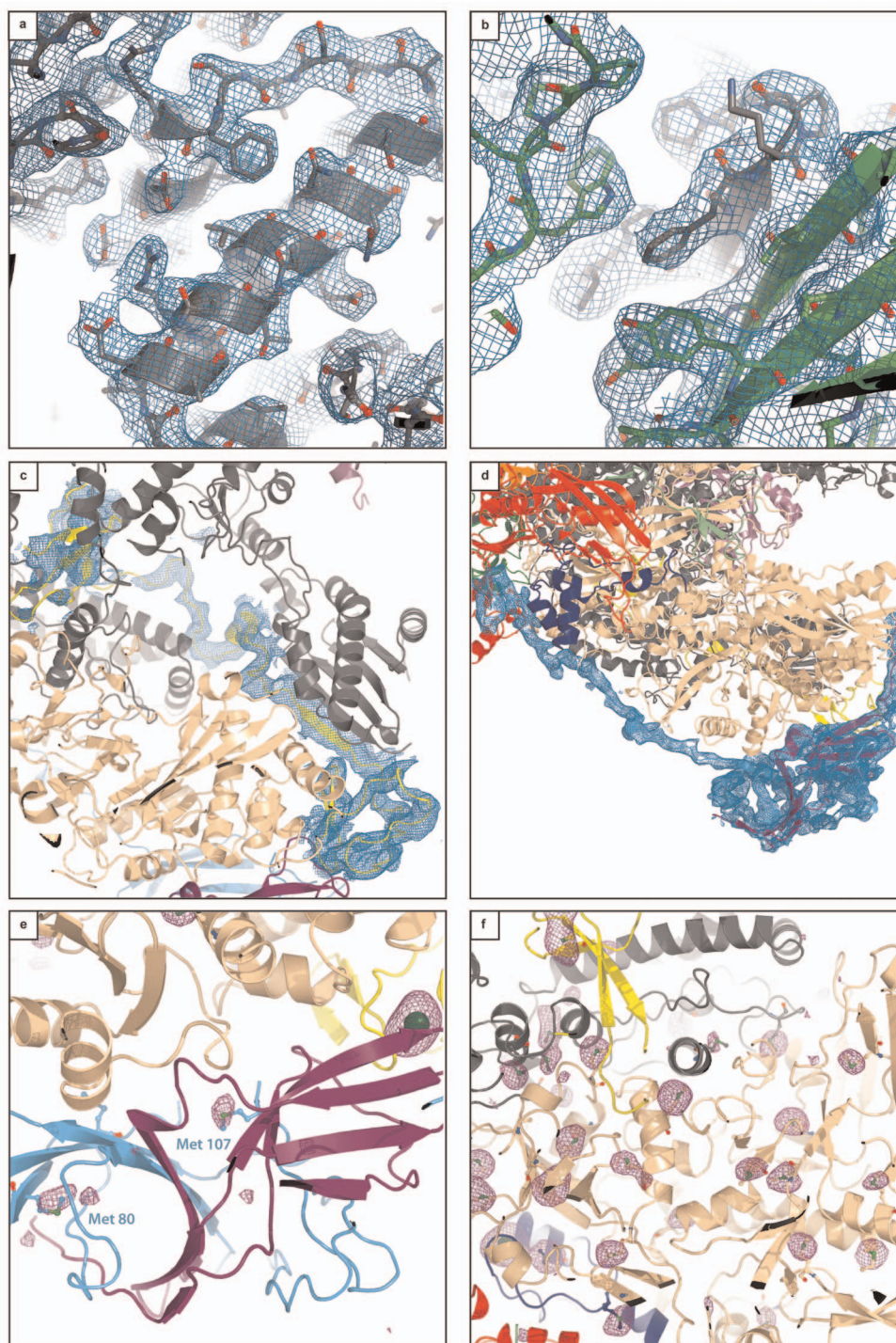
Data collection and structure determination. Data were collected at synchrotrons Soleil, ESRF, SLS and DESY and processed with XDS⁴⁵. All crystals belong to space group C2 but vary in cell dimensions (Extended Data Table 1). Crystal form C2-93 (β = 93°) was obtained in EG, whereas crystal forms C2-90 (β = 90°) and C2-100 (β = 100°) were grown in MPD. Anomalous data sets for tantalum bromide, selenomethionine and zinc were collected at the absorption edge of these elements using an inverse beam protocol. For the ytterbium derivative, the twofold symmetry axis of the crystal was aligned with the spindle axis to collect Bijvoet pairs on the same image and a multi-wavelength anomalous dispersion (MAD) data set was collected at four wavelengths. Initial phases were obtained in crystal form C2-93 through molecular replacement with Phaser⁴⁶ using the Pol II core structure, but did not allow further model building. After SIRAS phasing with SHARP⁴⁷ using a tantalum bromide derivative in this crystal form that diffracted to 6.5 Å and a native data set diffracting to 4.0 Å, followed by density modification and phase combination with the molecular replacement phases, an improved map was obtained. This allowed extension of the initial model with Coot⁴⁸ and subsequent refinement with Refmac⁴⁹. Better phases were obtained with a MAD experiment in crystal form C2-90 performed on an ytterbium derivative diffracting to 4.0 Å, and the model could be further improved in this space group and refined using Refmac. Initial phases in crystal form C2-93 were improved by MIRAS using a 3.6 Å native data set combined with tantalum bromide, ytterbium and lead derivative data sets, and combined with model phases followed by density modification. Multi-crystal averaging using improved native data sets diffracting to 3.3 Å in crystal forms C2-93 and C2-90 yielded maps of excellent quality where most side chains were visible, and thus the structure could be completed and further refined using Phenix⁵⁰. Finally, a data set was obtained in crystal form C2-100 and used for refinement with BUSTER⁵¹ against data diffracting to 3.03 Å. Further validation was performed with a data set of the selenium absorption spectrum that allowed locating 90 out of 100 selenium

atoms within a distance of less than 2.3 Å to corresponding methionine residues. Additionally, the position of the zinc atoms was also validated by inspection of anomalous difference maps. The final model in crystal form C2-100 has an *R*_{work}/*R*_{free} of 19.7/22.9% with excellent stereochemistry according to MolProbity⁵² (Extended Data Table 1) and only 7 residues in disallowed regions of the Ramachandran plot. Further experimental details on the X-ray structure determination will be published elsewhere.

Pol I RNA extension assay. RNA (5′-AAGACCAGGC-3′) was first radioactively labelled with ³²P by T4 PNK and gel purified on denaturing 15% urea-PAGE. After annealing of 2 pmol of radioactive RNA with 2 pmol of single-stranded DNA (template DNA: 3′-CCGTCATGATCATTACTGGTCCGCATTCATGA ACTCGAACC-5′ or 3′-GCTCAGCCTG GTCCGCATGTGTGCATGC-5′) the DNA/RNA hybrid was incubated with 4 pmol of Pol I for 20 min at 20 °C in 20 mM Tris pH 7.5, 150 mM KCl, 10 mM DTT and 10 mM MgCl₂. RNA elongation was initiated by addition of 1 mM NTPs in 60 mM ammonium sulphate, 20 mM Tris pH 7.5, 8 mM MgCl₂, 10 µM zinc chloride, 10% glycerol, and 10 mM DTT. After incubation for 20 min at 28 °C, RNA extension was stopped by addition of loading buffer (8 M urea, TBE) and heating for 2 min at 95 °C. The resulting P³²-labelled RNA product was subsequently analysed on a denaturing polyacrylamide gel (17% PAGE, 8 M urea). The gel was fixed in 10% ethanol, 6% acetic acid, dried and autoradiographed with an X-Ray film (Biomax, MR-film, Kodak).

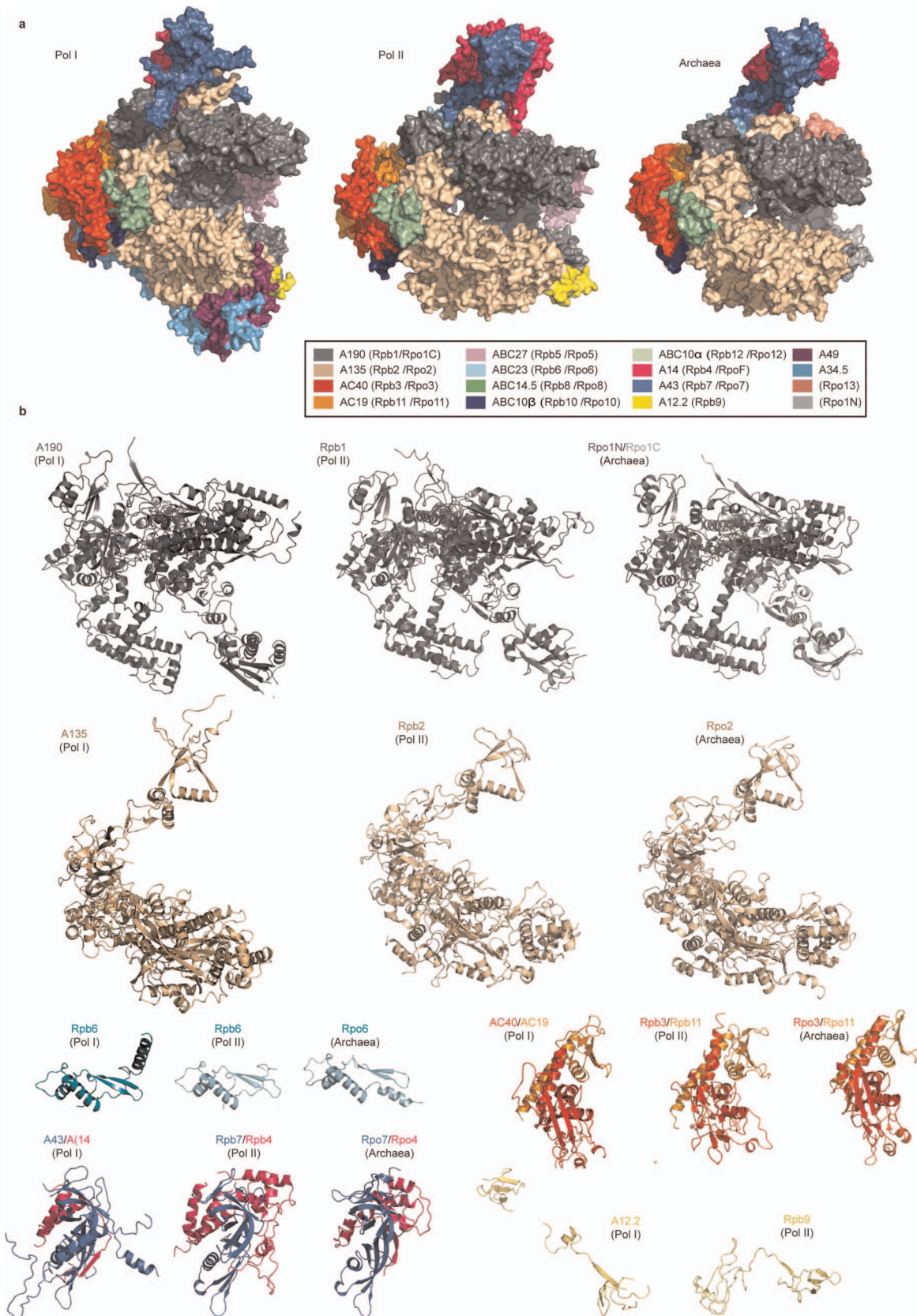
Construction of the yeast strain with DNA-mimicking loop deletion. To construct the yeast strain with a deletion of the DNA-mimicking loop in RPA190, a C-terminal fragment of RPA190 was amplified from pNOY16³³ using the primers RPA190-PstI-s (5′-GTCGTGAAGCCGCTGCAG-3′) and RPA190-STOP-EcoRI-as (5′-ACAGAATTCTGTGCTA GCCTAAGCCGCATTTGGAACCTTTG-3′) and cloned into PstI and EcoRI sites of pUC19, generating an NheI site after the RPA190 STOP codon. Subsequently, the plasmid without the DNA-mimicking loop, pUC19-rpa90Δloop, was generated by restriction-free cloning using the primers RPA190woloop-s (5′-GCTACAGACAGATGTTGCAATAGTTCTTTC GGAAATTGAACTATGAGAGAAGCTGAAAAGTCTTCT-3′) and RPA190 woloop-as (5′-AGAAGACTTTTCAGCTTCTCTCATAGTTTCAATTTCCGAA GAACATTTGCAACATCTGTCTGTAGC-3′). The NatNT2 marker was amplified from pFA6a-NatNT2⁵⁴ using the primers NatN2-NheI-s (5′-AAGCTAGC CGTACGCTGCAGGTCGAC-3′) and S2 (5′-ATCGATGAATTCGAGCTCG-3′) and cloned into the NheI and EcoRI sites of pUCrpa90Δloop to generate pUCrpa90Δloop-Nat. A deletion cassette was finally amplified from pUCrpa90Δloop-Nat using the primers loop-456-s (GGTCACGGTGCCGCTAATG) and LoopMarker as (5′-CCTTCTCCTTCAAATAAACTAATATTAATCGTAATAAATTATGG GACCTTTTGCTGCTTGGCCAGTGAATTCGAGCTCG-3′), purified and transformed into yeast strain SC1613 by the lithium acetate method⁵⁵. Colonies were selected on YPD plates containing 100 µg ml⁻¹ of ClonNat and analysed by colony PCR and subsequent sequencing. Growth of the yeast strain was monitored in comparison to the parental strain SC1613. Serial dilutions were done on SDC and YPD plates and incubated at 25 °C, 30 °C and 37 °C.

45. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
46. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
47. Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
49. Skubák, P., Murshudov, G. N. & Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D* **60**, 2196–2201 (2004).
50. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
51. Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68**, 368–380 (2012).
52. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
53. Wittekind, M. et al. Isolation and characterization of temperature-sensitive mutations in RPA190, the gene encoding the largest subunit of RNA polymerase I from *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 3997–4008 (1988).
54. Janke, C. et al. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).
55. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168 (1983).



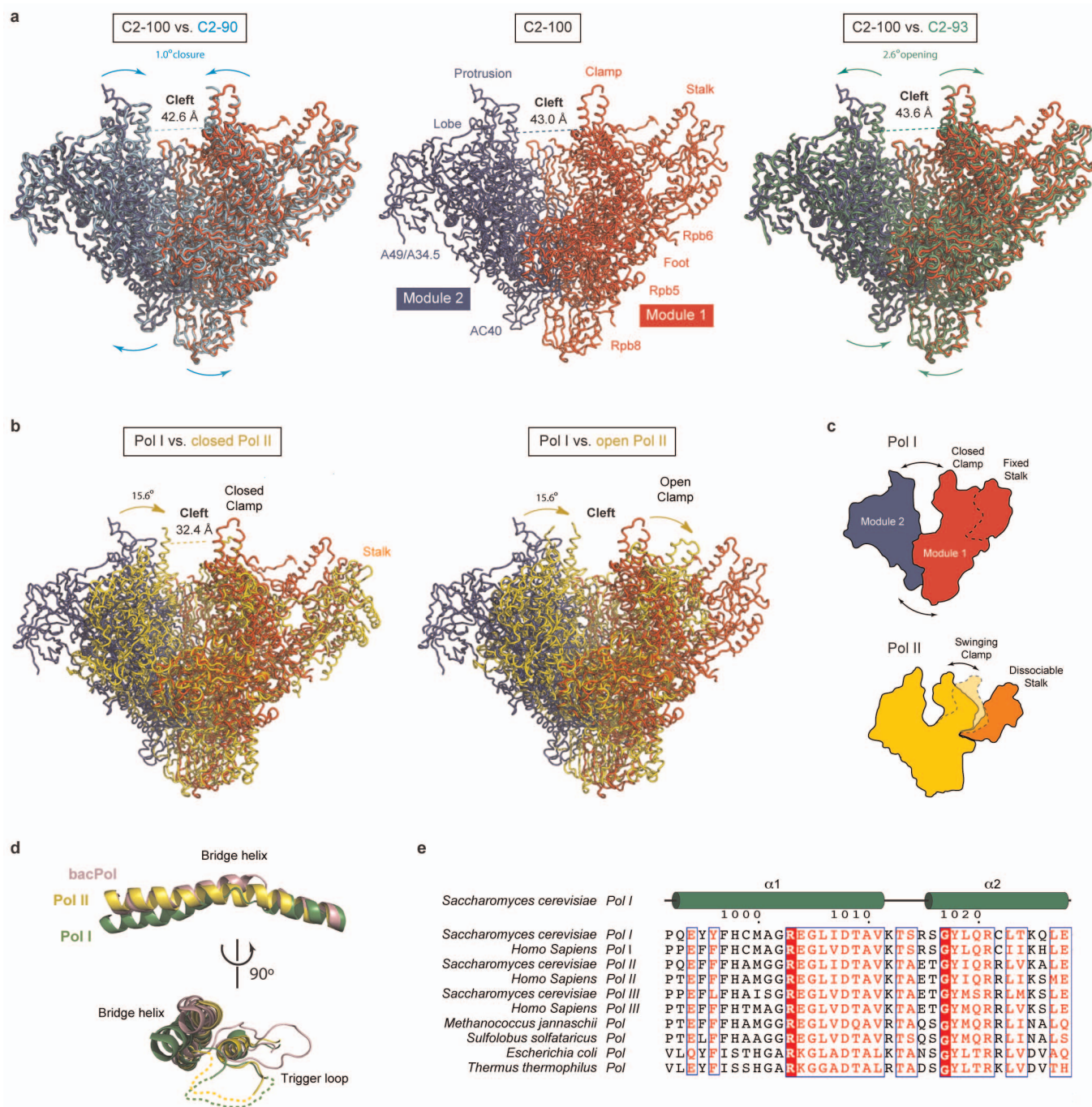
Extended Data Figure 1 | Electron densities of different regions of the Pol I structure. **a**, Helix in the cleft of subunit A190. Subunits are coloured according to the code given in Fig. 1a. In **a–d**, σ_A -weighted electron densities contoured at 1σ are depicted in blue. **b**, Interaction between subunit Rpb8 (green) and subunit A190 (grey). **c**, Linker between the two Zn ribbon domains in subunit A12.2. **d**, A49–A34.5 heterodimer and the anchoring onto the Pol I core by the A34.5 hook and the A49 linker region. **e**, Anomalous difference

Fourier map (purple) calculated from partially selenomethionine-substituted Pol I contoured at 3σ (Extended Data Table 1). In the A49–A34.5 heterodimer two selenium peaks correspond to A34.5 Met 80 and Met 107. **f**, Anomalous difference Fourier map (purple) showing selenomethionine positions contoured at 3σ . In total, 90 out of 100 expected selenium positions were located within a distance of less than 2.3 Å to corresponding methionine residues.



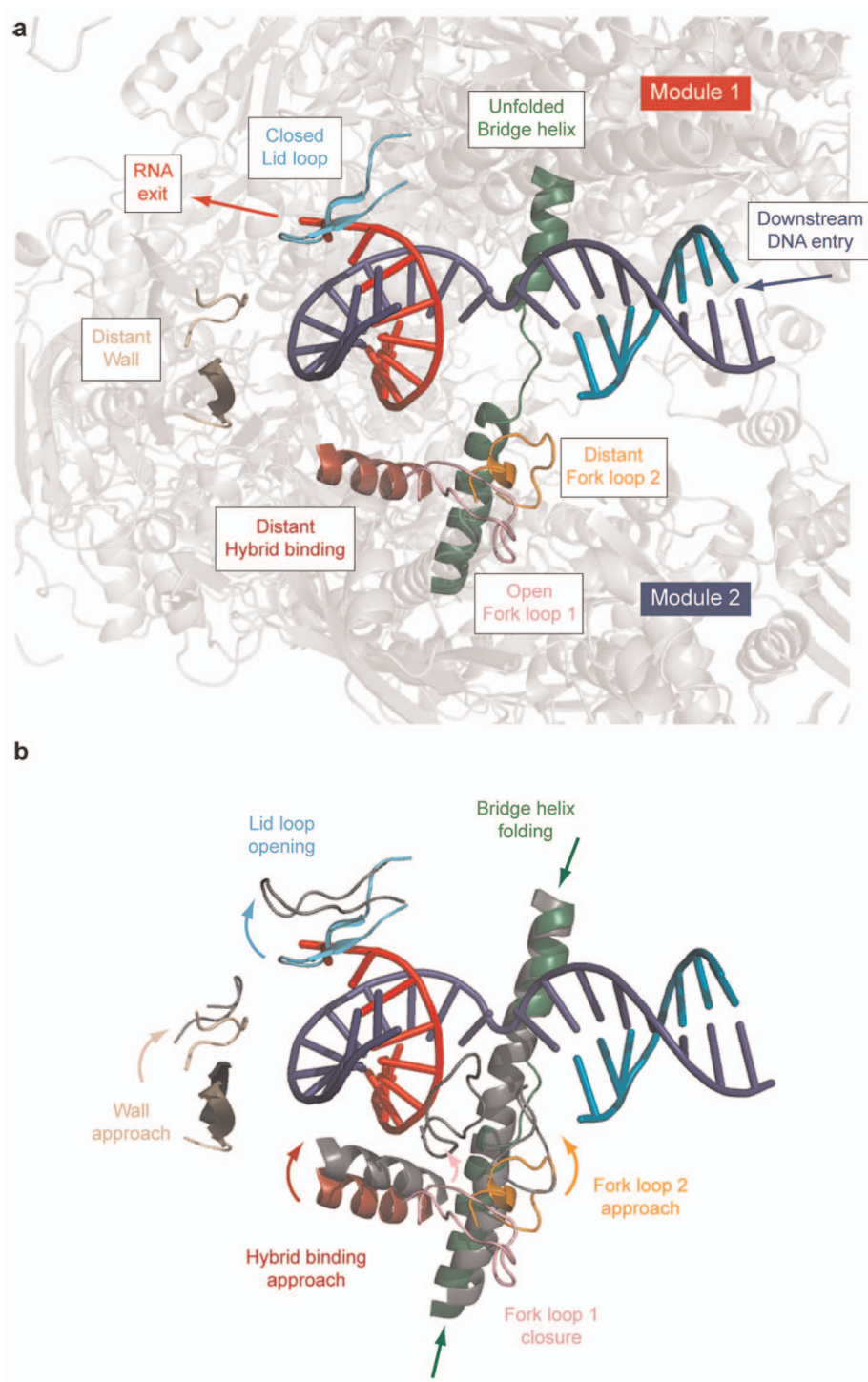
Extended Data Figure 2 | Comparison between Pol I, Pol II and archaeal Pol. **a**, Crystal structures of yeast Pol I and Pol II and the archaeal Pol are represented in the same orientation using the same colour code. Whereas the overall organization is conserved, additional subunits such as the A49–A34.5 heterodimer in Pol I and Rpo13 in Archaea are also present. The archaeal Pol

lacks the orthologue of subunit A12.2 in Pol I (or Rpb9 in Pol II). The relative position of the stalk also varies between the three RNA polymerases. **b**, Crystal structures of the individual subunits varying between the three enzymes are depicted. The same colour code for corresponding or identical subunits in the three RNA polymerases is used.



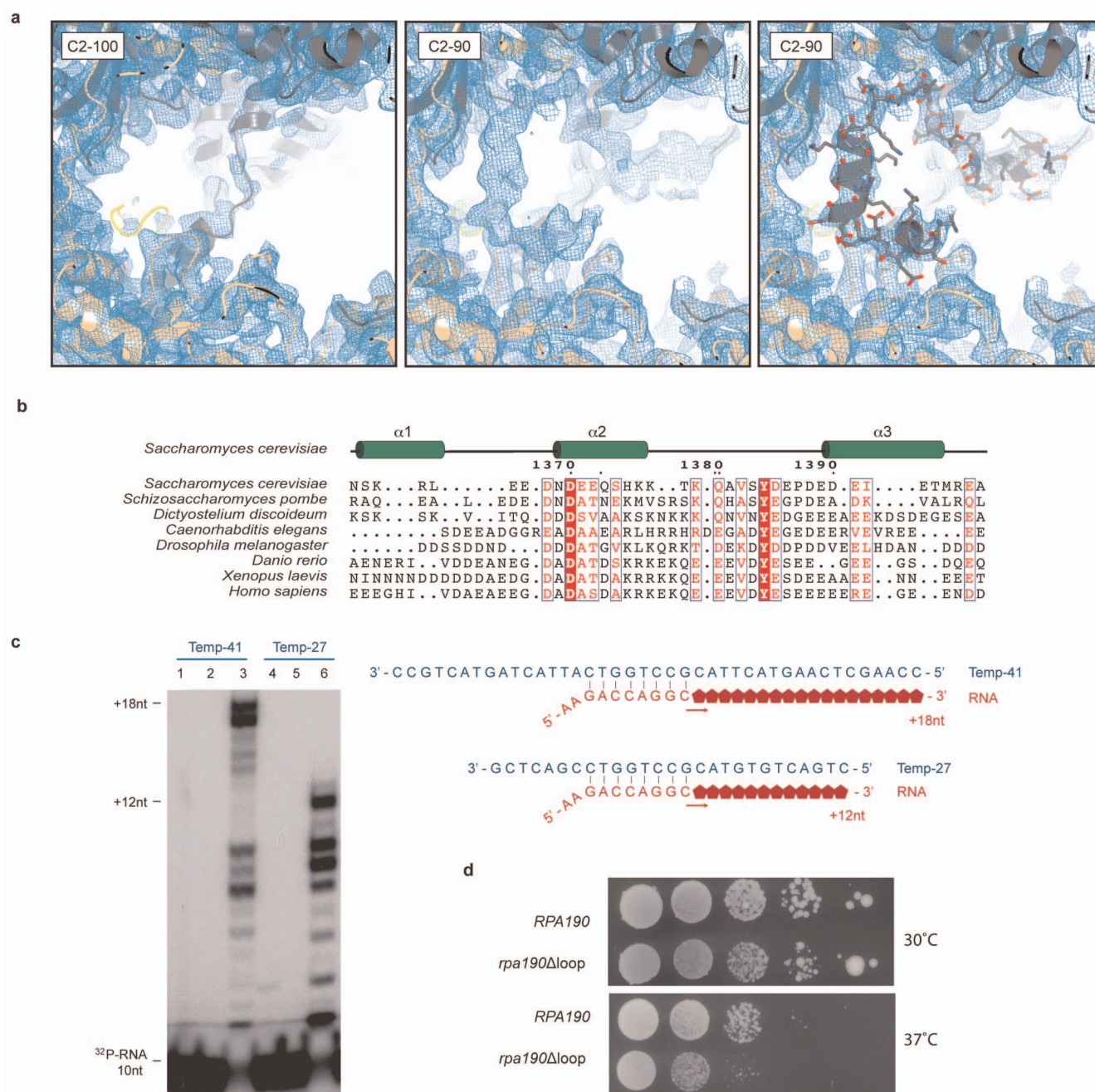
Extended Data Figure 3 | Opening of the cleft varies among different Pol I structures and between Pol I and Pol II. **a**, Middle panel: front view of the Pol I structure in crystal form C2-100 (see Extended Data Table 1). The complex is divided into two modules. Module 1 (red) is formed by the major part of subunit A190 (without the pore 1, funnel and jaw domains), the C terminus of A135, Rpb5, Rpb6, Rpb8 and the stalk subunits, whereas module 2 (blue) comprises the remaining A135 domains, the pore 1, funnel and jaw domains of A190, AC40–AC19, Rpb10, Rpb12, A12.2 and the A49–A34.5 heterodimer. These modules are held by three hinges in A190 (active site–pore 1 connection, bridge helix and jaw–cleft connection) and one hinge in A135 (hybrid binding–anchor connection) as indicated in Fig. 1b. Pol I structures obtained in crystal forms C2-90 (left panel) and C2-93 (right panel) were superimposed with the one obtained in crystal form C2-100 taking module 2 as reference. Differences between the different crystal forms in the cleft aperture and the tilting of the mobile modules are indicated. **b**, Pol II structure (Protein Data Bank accession 1WCM) is superimposed onto Pol I (C2-100) using module 1 as reference. In

comparison, the Pol II cleft is closed by 10 Å and the modules rotate 15.6°. **c**, Schematic representation of Pol I and Pol II showing the mobility between Pol I modules, as well as the conformation of the stalk and the clamp. Colour coding is as in **b**, with the exception of the Pol II stalk, which is coloured deep orange. **d**, Conformation of the bridge helix of the bacterial *Thermus thermophilus* polymerase (bacPol, pink, Protein Data Bank accession 1IW7), Pol I (green) and Pol II (yellow, Protein Data Bank accession 1WCM). In addition, the trigger loop is shown for the *Thermus thermophilus* polymerase, where it is ordered, and as dotted lines for Pol I and Pol II, where it is disordered. **e**, Sequence alignment of the bridge helix of *Saccharomyces cerevisiae* and *Homo sapiens* Pol I, Pol II and Pol III, archaeal *Methanococcus jannaschii* and *Sulfolobus solfataricus*, and the bacterial *Escherichia coli* and *Thermus thermophilus* polymerases. The secondary structure of the Pol I bridge helix is shown above the alignment. In *Methanococcus jannaschii*, site-directed mutations Q823D and S824P in subunit A' lead to increased transcriptional activity.



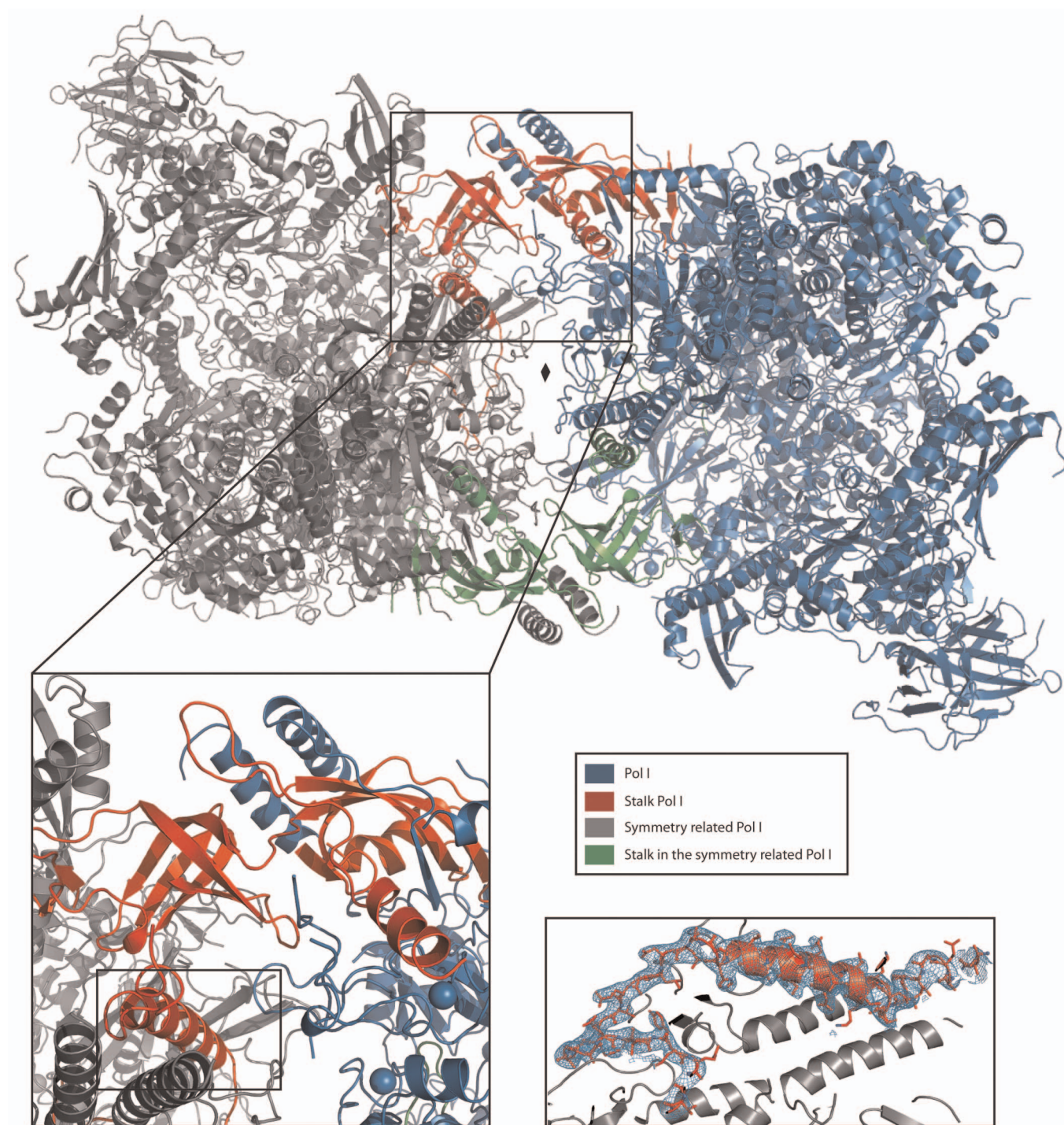
Extended Data Figure 4 | Elongation complex. **a**, Cartoon representation of a model of Pol I in complex with an elongation bubble, generated by superposition of the Pol II elongation complex crystal structure (Protein Data Bank accession 1Y1W) using the largest subunit as reference. Whereas Pol II is not shown, the coding and non-coding DNA strands are depicted in blue and cyan, respectively, and the RNA in red. The main Pol I elements putatively involved in nucleic acid interaction appear in different colours, whereas the rest

of the Pol I structure is shown in light grey. **b**, Proposed rearrangements in elongating Pol I (coloured elements) in analogy with Pol II (grey elements). Closure of the cleft is expected to approach the wall (tan), the hybrid binding domain (dark red) and the fork loop 2 (orange) to the bubble, as well as to fold the bridge helix (green). Closure of the fork loop 1 (pink) and opening of the lid loop (cyan) would also be required.



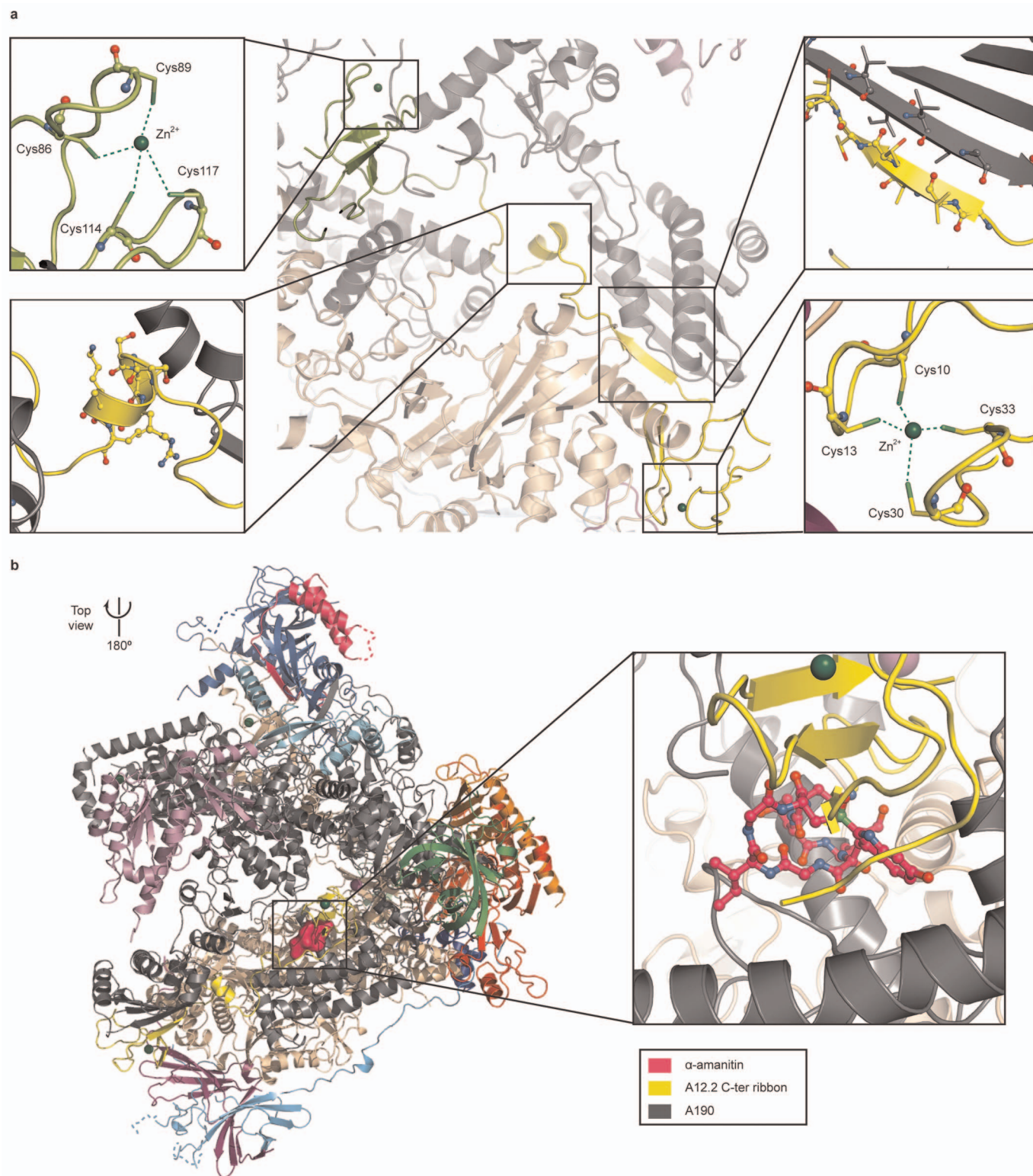
Extended Data Figure 5 | The DNA-mimicking loop of Pol I forms a mobile element. **a**, σ_A -weighted electron density contoured at 1σ of A190 jaw residues 1361–1399 (middle and right panel) in crystal form C2-90. Density is also present in crystal form C2-93 (data not shown), whereas it is absent in crystal form C2-100 at the same contour level (left panel). **b**, Sequence alignment of the Pol I DNA-mimicking loop across different species highlighting the conservation of this element. **c**, Purified Pol I shows elongation activity in an RNA extension assay. DNA templates (Temp-41 and Temp-27 of 41 and 27 nucleotides, respectively) and ^{32}P -labelled RNA sequences used for the assay

are indicated. The autoradiogram shows the elongation of RNA by Pol I producing a run-off of 18 nucleotides (lane 3) or 12 nucleotides (lane 6) depending of the template used. Lanes 1 and 4: the DNA/RNA hybrids were incubated in the absence of Pol I. Lanes 2 and 5: Pol I–DNA–RNA complexes were incubated with a buffer without NTPs. **d**, Dot spots grown at the indicated temperatures of the parental *RPA190* strain and *rpa190* Δ loop strain where the DNA-mimicking loop has been deleted. The *rpa190* Δ loop strain shows a slight temperature-sensitive growth defect on SDC medium.



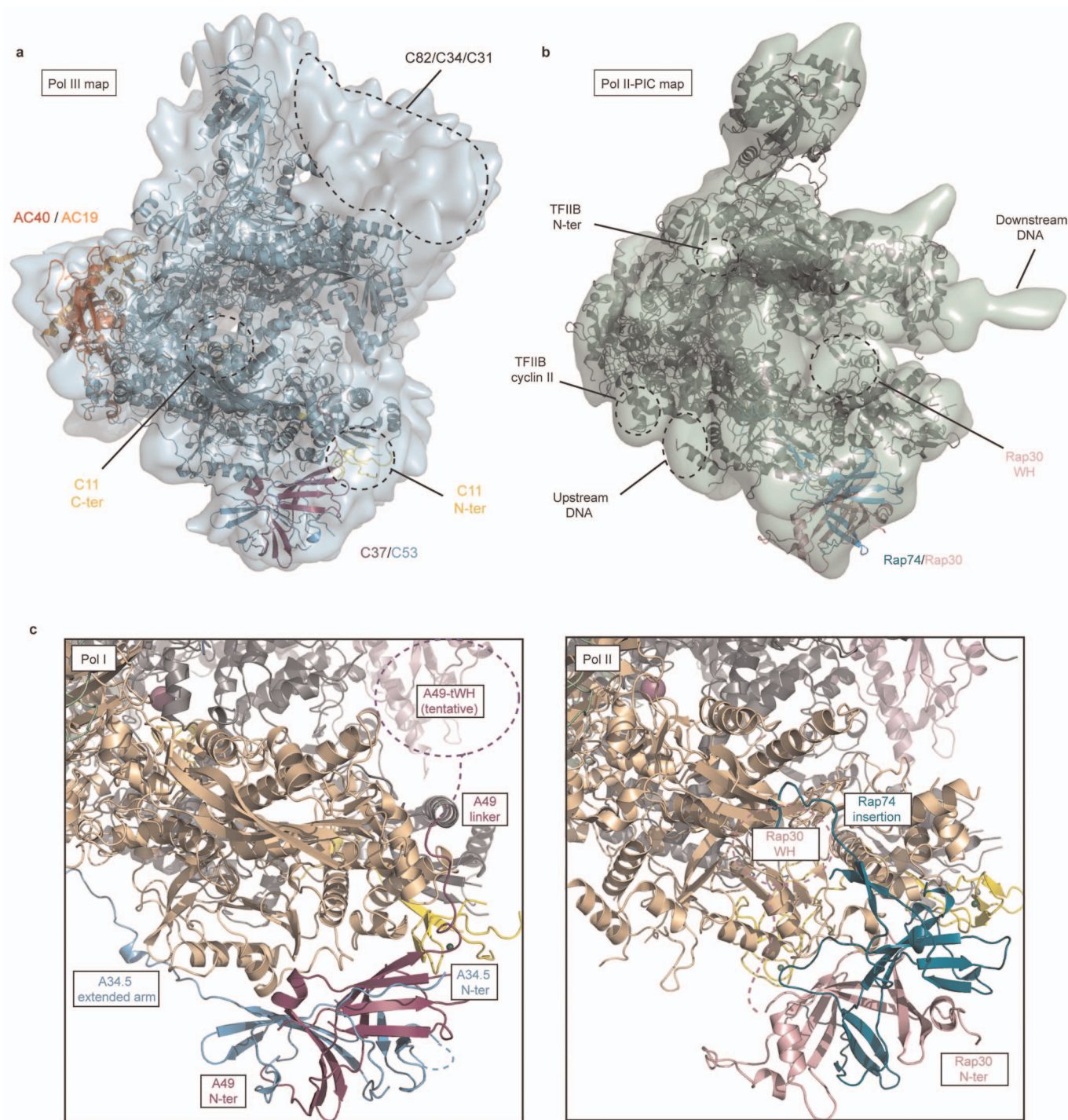
Extended Data Figure 6 | Pol I dimer in the crystal lattice. The A43 C-terminal tail establishes crystal contacts with a second molecule related by a crystallographic dyad. The A43 C-terminal helix is embedded between the clamp and the protrusion domain of a dyad related molecule. The σ_A -weighted

electron density map (contoured at 1σ) shows clear density corresponding to residues A43 251–326. The two monomers are related by a crystallographic dyad, which is indicated by a dyad symbol.



Extended Data Figure 7 | Subunit A12.2 structure and its position in Pol I.
a, Detailed views show the A12.2 Zn sites and the main contacts between its linker and the A190 subunit. The A12.2 linker extends the β -sheet of the A190 jaw. **b**, The overlap between subunit A12.2 and α -amanitin in the Pol I structure explains the insensitivity of Pol I for this fungal toxin. The Pol II- α -amanitin

complex structure (Protein Data Bank accession 2VUM) was superimposed onto the Pol I crystal structure. In the left panel, the α -amanitin toxin is depicted in surface representation (pink). On the right, a detailed view of α -amanitin shows the overlap with the C-terminal Zn ribbon of A12.2.



Extended Data Figure 8 | Precise positioning of the A49–A34.5 heterodimer suggests similar positions for the related C37–C53 heterodimer in Pol III and the TFIIIF heterodimer in Pol II. **a**, Pol I was fitted into the Pol III envelope (EM-1804)¹². A49–A34.5 (corresponding to C37–C53 in Pol III), AC40–AC19 and A12.2 (corresponding to C11 in Pol III) are coloured as in Fig. 1. The approximate position of subcomplex C82–C34–C31 is also indicated. **b**, The proposed Pol II/TFIIIF model was manually fitted into

the Pol II/TFIIIF-A-B-F/TBP/DNA EM density (EM-2305)⁴³. **c**, Left panel: detailed view of the anchoring of the A49–A34.5 dimerization domain onto the Pol I core. Right panel: model for the TFIIIF dimerization module bound to the Pol II core based on the crystal structures of the human Rap74–Rap30 complex (Protein Data Bank accession 1F3U) and Pol II (Protein Data Bank accession 1WCM).

Extended Data Table 1 | Data statistics of the Pol I structure determination

	C2-90 SeMet	C2-90	C2-93	C2-100
Data collection				
Space group	C2	C2	C2	C2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	403.5, 140.4, 142.1	401.0, 139.8, 140.9	425.2, 140.6, 139.7	400.5, 140.2, 122.9
β (°)	90.0	90.4	93.4	100.1
Beamline	SOLEIL (Proxima 1)	SOLEIL (Proxima 1)	SOLEIL (Proxima 1)	DESY PETRA III (P14)
Resolution (Å)	71–3.59 (3.69–3.59)*	80–3.35 (3.44–3.35)*	90–3.27 (3.35–3.27)*	84–3.03 (3.11–3.03)*
$R_{\text{merge}}^{\dagger}$	0.192 (3.711)	0.193 (2.163)	0.162 (5.838)	0.104 (2.322)
$I/\sigma I$	11.2 (0.9)	8.5 (0.9)	10.9 (0.5)	12.3 (0.8)
Completeness (%)	99.6 (95.2)	100 (100)	100 (99.9)	99.1 (99.2)
Redundancy	14.3 (9.8)	7.7 (7.9)	10.9 (11.2)	7.1 (7.1)
CC(1/2) [‡]	0.999 (0.375)	0.997 (0.429)	0.999 (0.317)	0.999 (0.336)
Refinement				
PDB ID		4C3J	4C3H	4C3I
Resolution (Å)	N/A§	50–3.35	50–3.27	84–3.03
No. reflections		105911	124088	128861
$R_{\text{work}}/R_{\text{free}}$		21.3/22.0	22.1/22.6	19.7/22.9
No. atoms				
Protein		34545	34545	34221
Ligand/ion		7 (Zn)	7 (Zn)	31 (including 7 Zn)
Water		0	0	0
B-factors				
Protein		135.1	171.9	119.9
Ligand/ion		128.3	160.4	134.1
R.m.s deviations				
Bond lengths (Å)		0.008	0.008	0.008
Bond angles (°)		0.96	0.96	0.96

* Highest resolution shell is shown in parentheses.

[†] Data were collected from one crystal.[‡] According to ref. 18. For C2-100 the highest resolution is 3.15 Å and 3.27 Å resolution using the criteria $CC_{1/2} > 0.5$ and $I/\sigma I > 2.0$, respectively.[§] Data set C2-90 SeMet was only used for anomalous difference map, not for model building.

RNA polymerase I structure and transcription regulation

Christoph Engel¹, Sarah Sainsbury¹, Alan C. Cheung¹, Dirk Kostrewa¹ & Patrick Cramer¹

Transcription of ribosomal RNA by RNA polymerase (Pol) I initiates ribosome biogenesis and regulates eukaryotic cell growth. The crystal structure of Pol I from the yeast *Saccharomyces cerevisiae* at 2.8 Å resolution reveals all 14 subunits of the 590-kilodalton enzyme, and shows differences to Pol II. An ‘expander’ element occupies the DNA template site and stabilizes an expanded active centre cleft with an unwound bridge helix. A ‘connector’ element invades the cleft of an adjacent polymerase and stabilizes an inactive polymerase dimer. The connector and expander must detach during Pol I activation to enable transcription initiation and cleft contraction by convergent movement of the polymerase ‘core’ and ‘shelf’ modules. Conversion between an inactive expanded and an active contracted polymerase state may generally underlie transcription. Regulatory factors can modulate the core–shelf interface that includes a ‘composite’ active site for RNA chain initiation, elongation, proofreading and termination.

Eukaryotic genes are transcribed by Pol I, II and III (ref. 1), which synthesize ribosomal, messenger and transfer RNA, respectively. Pol I synthesizes about 60% of cellular RNA by transcribing several copies of the rRNA gene. Each gene copy is transcribed by dozens of Pol I enzymes, generating microscopic ‘Christmas tree’ structures². Pol I activity is a key determinant for the level of ribosome components and controls cell growth³. Pol I deregulation occurs in many diseases including cancer⁴.

Two decades ago the shape of Pol I was revealed by electron microscopy^{5,6}, but the crystal structure of Pol I is lacking. The structure of Pol II was, however, solved for its 10-subunit core at the turn of the millennium⁷. Subsequent structures of Pol II complexes have elucidated the transcription mechanism⁸. Pol II contains a conserved active centre cleft with a catalytic aspartate loop, a bridge helix, and a clamp that closes over the DNA–RNA hybrid during transcription. Pol I contains a Pol-II-like core consisting of five homologous (A190, A135, AC40, AC19 and A12.2) and five common (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12) subunits, and the heterodimeric subcomplexes A14–A43 and A49–A34.5 (Extended Data Table 1).

To determine the Pol I structure, we previously established large-scale purification of the endogenous enzyme from the yeast *S. cerevisiae*⁹. We obtained crystals of Pol I in 2005, but these diffracted only to low resolution (C. D. Kuhn and P.C., unpublished observations). We therefore studied Pol I by electron microscopy, homology modelling and X-ray analysis of its A14–A43 subcomplex⁹. We showed that the A49–A34.5 subcomplex contains an amino-terminal dimerization domain and a carboxy-terminal tandem winged-helix (tWH) domain in A49, and solved the crystal structures of both domains¹⁰. We then used crosslinking to place A14–A43 and A49–A34.5 on the surface of a Pol I core model¹¹. Here we report the crystal structure of Pol I, which reveals new polymerase elements and provides unexpected insights into transcription regulation.

Pol I structure

In a long-standing search for well-ordered Pol I crystals, we changed parameters in yeast fermentation, protein purification and crystallization, and eventually obtained a crystal form amenable for structure determination (Methods). Purified Pol I was active in DNA-templated

RNA extension and in RNA cleavage (Extended Data Fig. 1a). The structure was solved with phases from intrinsic zinc ions and refined at 2.8 Å resolution to an R_{free} factor of 21.0% (Extended Data Table 2a, b). The structure comprises a Pol I dimer containing 8,681 amino acid residues, and lacks only the mobile A49 tWH domain and several surface loops.

The structure reveals the 10-subunit Pol I core and the subcomplexes A49–A34.5 and A14–A43 on opposite sides (Fig. 1). Compared with the largest Pol II subunit Rpb1, subunit A190 contains insertions in its domains clamp, dock, cleft, pore and funnel, and lacks parts of the clamp head, foot, jaw and cleft domains (Figs 1c and 2a, c and Extended Data Figs 2 and 4a). The jaw domain contains a long insertion that we name ‘expander’. The clamp head contains a helical insertion at the position of the archaeal polymerase subunit Rpo13 (ref. 12). Compared to the Pol II subunit Rpb2, subunit A135 contains a truncated external 1 domain, an extended protrusion, and a clamp insertion (Figs 1c and 2b, d and Extended Data Figs 3 and 4b). A heterodimer of AC40 and AC19 resembles Rpb3–Rpb11 in Pol II, but AC40 contains an additional ‘toe’ domain where the archaeal subunit D contains a domain with an iron-sulphur cluster¹³ (Fig. 1c and Extended Data Fig. 5). Subunit A12.2 binds with its N-terminal zinc ribbon domain (N-ribbon) where the homologous Rpb9 domain binds Pol II (ref. 7), but its C-ribbon binds in the pore like transcription elongation factor IIS (TFIIS)¹⁴ (Fig. 3a, b and Extended Data Fig. 6a), consistent with crosslinking¹¹.

Composite active centre

Whereas Pol II contains a ‘tunable’ active site that is transiently complemented by the C-ribbon in TFIIS¹⁴, Pol I contains a ‘composite’ active site that comprises the A12.2 C-ribbon. The C-ribbon occludes the binding site for α -amanitin, explaining why Pol I is less sensitive to this toxin than Pol II (ref. 15). The A12.2 C-ribbon contains a hairpin that reaches the active site like the catalytic hairpin of TFIIS, explaining the strong RNA cleavage activity of Pol I (ref. 9). Pol III contains a corresponding C-ribbon in its subunit C11 (ref. 16). The composite active site of Pol I and III enables efficient proofreading^{9,17} and termination^{18,19}.

The A49–A34.5 subcomplex stabilizes A12.2 (Fig. 3c), explaining why it stimulates A12.2-dependent⁹ RNA cleavage¹⁰, and why it

¹Gene Center and Department of Biochemistry, Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany.

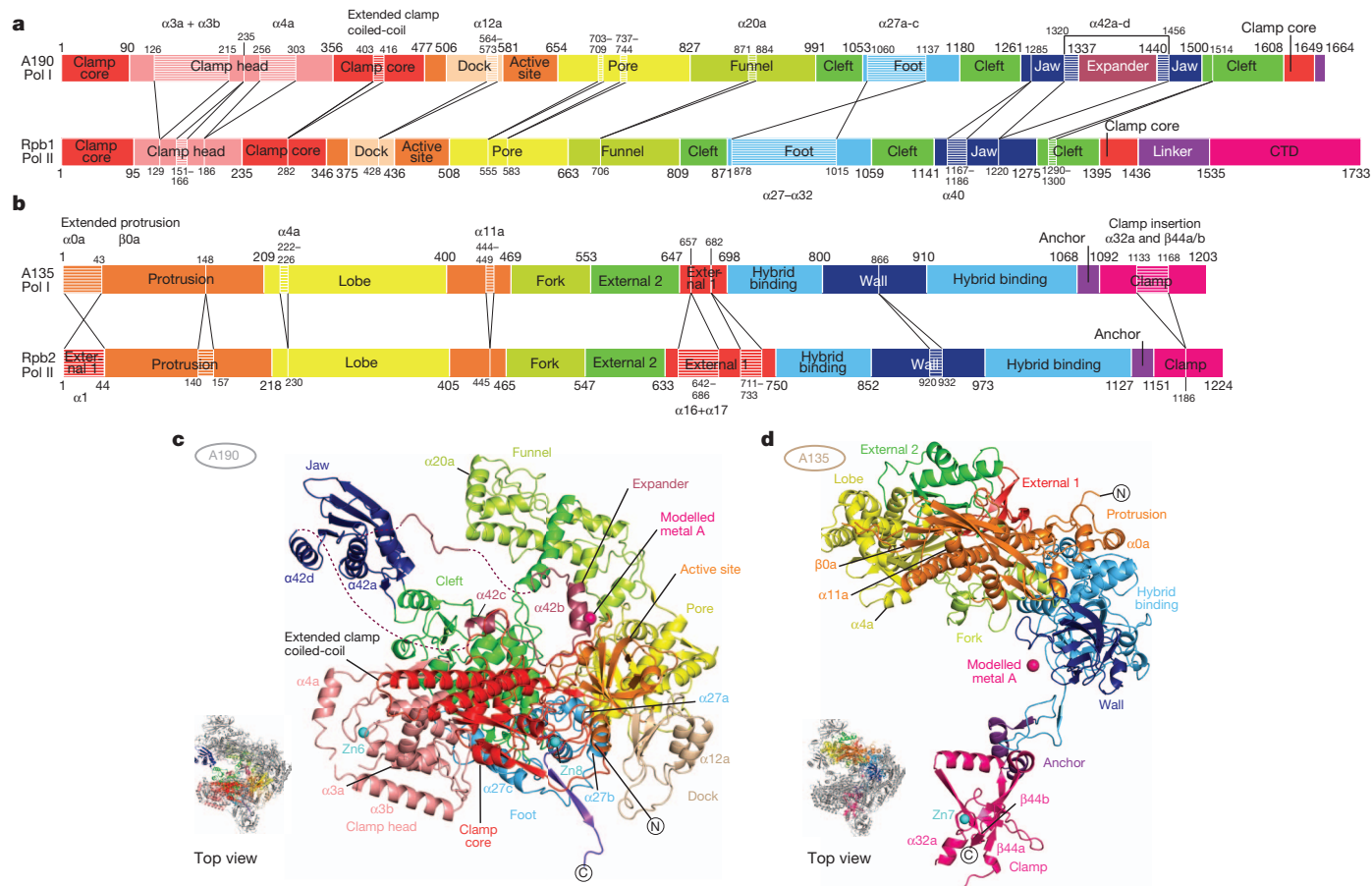
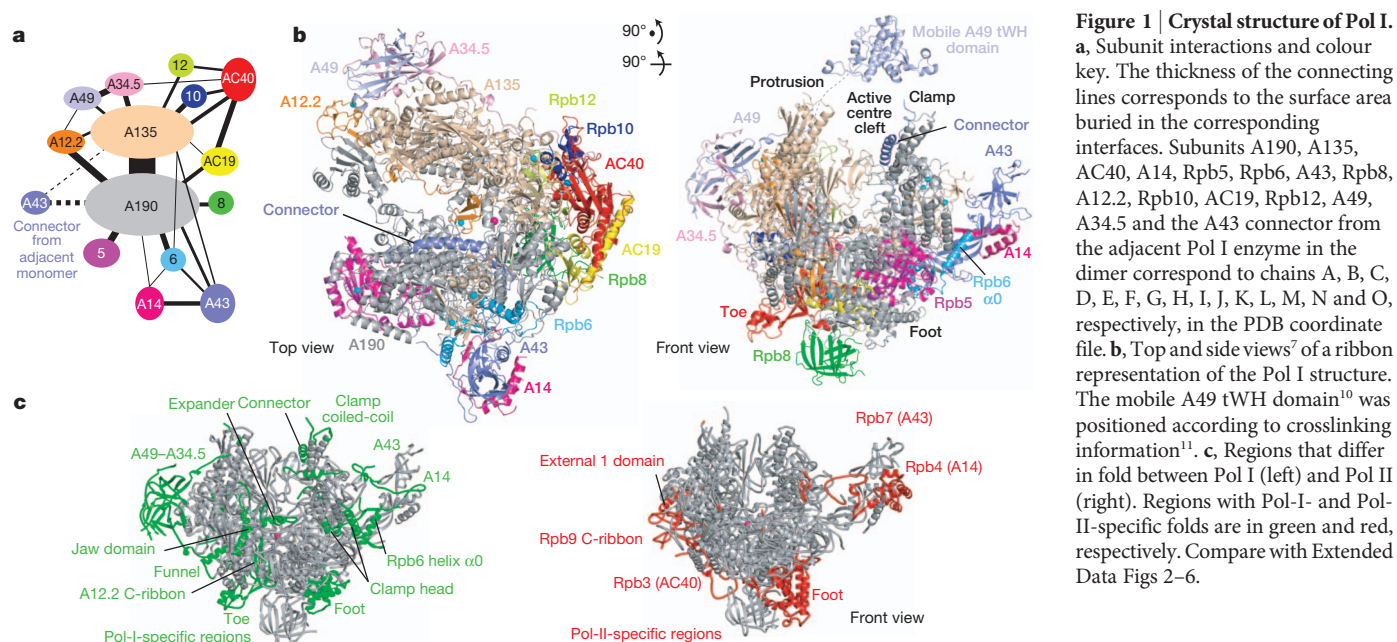


Figure 2 | Structure of the two largest Pol I subunits. **a**, Schematic of domains and domain-like regions of the largest subunit A190, based on the Pol II nomenclature⁷. The amino acid residue numbers at the domain boundaries are indicated. Apparent amino acid insertions and deletions compared with Rpb1 are marked. **b**, Schematic of domains and domain-like regions of the second largest Pol I subunit A135, based on the Pol II nomenclature⁷ as in **a**. **c**, Ribbon diagram of A190. The thumbnail shows the location of A190 within Pol I. Locations of N and C termini are indicated. Colour-coding as in **a**. Top view. **d**, Ribbon diagram of A135 as in **c**. The thumbnail shows the location of A135 within Pol I. See Extended Data Fig. 3 for further details. Top view.

view as in Fig. 1. Labelling of corresponding secondary structure elements is as for Pol II (ref. 7). New or lacking secondary structure elements are labelled. New elements were named according to the preceding Pol II element, with lower-case letters added alphabetically for subsequent elements. See Extended Data Fig. 2 for further details. **d**, Ribbon diagram of A135 as in **c**. The thumbnail shows the location of A135 within Pol I. See Extended Data Fig. 3 for further details.

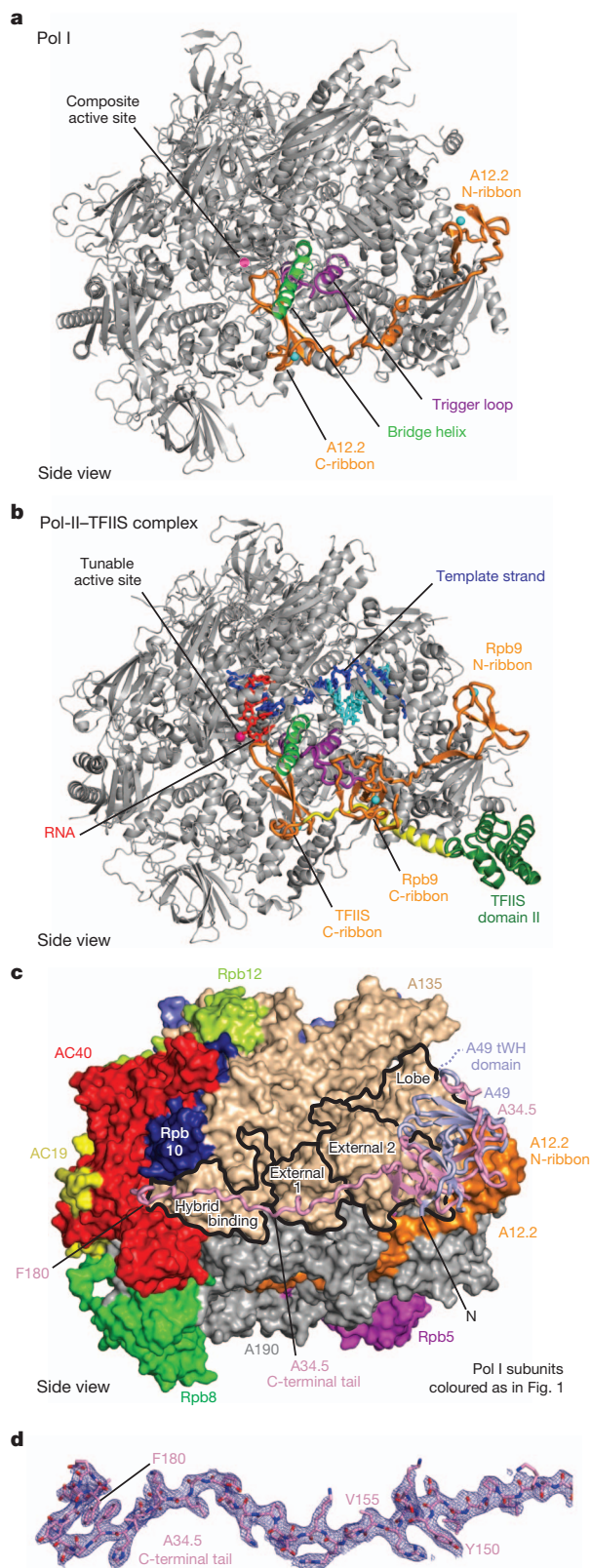


Figure 3 | Subunit A12.2 and the TFIIF-like subcomplex A49–A34.5. **a**, Location of A12.2 (orange) within Pol I (silver) viewed from the side. The location of the N- and C-ribbon domains is indicated. **b**, Structures of Rpb9 (orange) and TFIIS (orange, yellow, green) within Pol II¹⁴ (silver), viewed as in **a**. **c**, Ribbon model of A49–A34.5 on the Pol I core surface viewed from the side. On the basis of a comparison with the Pol II-TFIIF complex^{22,23}, A49 and A34.5 correspond to TFIIF subunits Tfg1 and Tfg2, respectively. The jaw-lobe module is stabilized by the A49–A34.5 subcomplex in a position close to the protrusion. **d**, $2F_o - F_c$ electron density (blue mesh, contoured at 1σ) for the positively charged A34.5 C-terminal tail that contributes to A49–A34.5 binding¹⁰.

dissociates after A12.2 deletion²⁰. A49–A34.5 stretches over the polymerase surface, to reach the AC40 toe domain with its positively charged A34.5 C-terminal tail (Fig. 3c, d and Extended Data Fig. 6b). The mobile A49 tWH domain (Fig. 1b) may close over nucleic acids in the cleft¹¹ to increase transcription processivity^{9,21}. The A49–A34.5 subcomplex is related to the Pol II initiation factor TFIIF and to the Pol III subcomplex C37–C53, which contain similar dimerization domains located at corresponding positions^{22–24}.

Expander and cleft expansion

Compared to Pol II, the active centre cleft of Pol I is expanded by 8 Å (Fig. 4 and Supplementary Video). This results from an apparent movement of the two major polymerase modules⁷, ‘core’ and ‘shelf’, which mainly comprise subunits A135 and A190, respectively (Fig. 4a, b and Extended Data Fig. 7). Cleft expansion differs from cleft opening by swinging of the clamp⁷, and includes a relative rotation of the core and shelf modules around an axis through the active site, similar to the ratcheting observed for bacterial RNA polymerase bound by an inhibitor protein²⁵ and in a paused state²⁶. The expanded cleft is stabilized by the expander, which forms two α -helices in the active centre (Fig. 4c). The expander binds the bridge helix, which is unwound in its central region (Fig. 4d). Expander residues Asp 1385 and Asp 1388 sandwich the Pol I-specific residue Arg 1015 in the unwound bridge helix (Fig. 4e). The expander residue Tyr 1384 inserts between the core and shelf modules, whereas Lys 1377 reaches the aspartate loop (Fig. 4e).

Cleft expansion changes the active centre that lies at the core-shelf interface (Fig. 5a). A tilting between the A190 active site domain and the A135 hybrid-binding domain alters the conformation of the catalytic aspartate loop and loop β 26– β 27 that emanate from these domains (Fig. 5b, c). In the tilted conformation, binding of catalytic metal ions and the RNA 3' end as in Pol II (refs 7, 27, 28) is impossible. Cleft contraction may rearrange the active site such that the aspartate loop can bind metal ions with its residues Asp 627, Asp 629 and Asp 631, and the A135 loop β 26– β 27 can bind the RNA 3' end with its residues Lys 916 and Lys 924. Binding of the substrate triphosphate may differ in Pol I because the A12.2 hairpin shields the A135 residues Arg 714 and Arg 957, the counterparts of triphosphate-binding Pol II residues. The A12.2 hairpin apparently interferes with closure of the triphosphate-binding trigger loop^{27,28}, but may bind the triphosphate and position metal ion B, as suggested for the TFIIS hairpin during RNA cleavage¹⁴.

Connector and polymerase dimerization

The observed dimeric form of Pol I is established by mutual interactions of the expanded cleft of one polymerase with the stalk subcomplex A14–A43 (refs 9, 29) of the other polymerase, resulting in a handshake (Fig. 6). A14–A43 binds to the core enzyme similar to its counterparts Rpb4–Rpb7 in Pol II (ref. 30) and C17–C25 in Pol III (refs 24, 31). A43 also binds a previously unobserved N-terminal α -helix in Rpb6 (Fig. 1b), explaining the known interaction between A43 and Rpb6 (ref. 29), and loss of A43 and Rpb6 from Pol I after A14 deletion³². Stalk-mediated Pol I dimerization was also observed by electron microscopy⁶, suggesting that the dimer observed in the crystal corresponds to that formed in solution.

The flexibly linked C-terminal region of A43 folds into a ‘connector’ that invades the cleft of the adjacent polymerase (Fig. 6a, b and Extended Data Fig. 8a). The connector forms a helix (K5) that binds along the clamp coiled-coil, a β -hairpin (D1–D2) that binds the RNA exit path and the lid, and an acidic C-terminal tail that binds the polymerase switch regions 2 and 3 (Fig. 6a). The connector buries 2,350 Å² of surface area between the two polymerases, consistent with a role in dimer stabilization. The defined structure of the connector and its specific interactions with the polymerase cleft argue that connector-mediated polymerase dimerization is functionally significant. Consistent with this, the connector cannot be accommodated in a

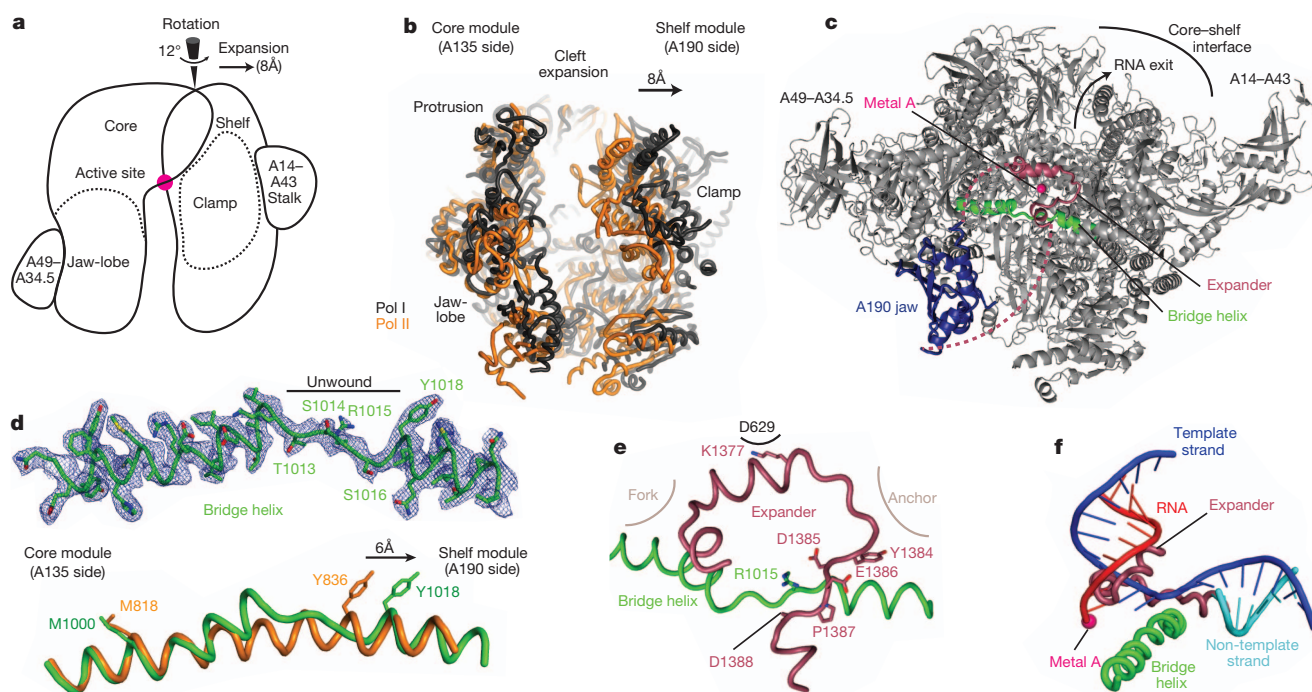


Figure 4 | Expanded cleft, expander and unwound bridge helix. **a**, Schematic of Pol I structure indicating the four mobile modules (core, shelf, clamp and jaw-lobe) and the two peripheral subcomplexes. The view is from the top⁷ as in Fig. 1, but rotated anticlockwise by 90°. The direction of cleft expansion by movement of the shelf module away from the core module is indicated by an arrow, and goes along with a 12° rotation between the two modules around the indicated axis through the active site. **b**, Expansion of the Pol I active centre cleft. Pol I (black) and Pol II (orange) were superimposed with their subunits A135 and Rpb2, respectively (left side). The view is as in **a**. **c**, A novel expander element (raspberry) that is inserted into the A190 jaw domain (blue) stabilizes the expanded Pol I cleft and a stretched and unwound bridge helix (green). **d**, $2F_o - F_c$ electron density (blue mesh, contoured at 1.3σ) for the bridge helix (green) reveals a central, unwound region. At the bottom, a comparison of the bridge helix in Pol II (orange), aligned on the core module side, reveals an expansion by 6 Å (arrow). View from the front as in Fig. 1. **e**, The expander (raspberry) stabilizes the unwound central bridge helix (green). View as in **c**. **f**, Binding of the expander and the DNA-RNA hybrid in the active centre are mutually exclusive. Superposition of the Pol I structure with the Pol II elongation complex structure (PDB code 1Y1W) reveals clashes between the expander (raspberry) and the DNA template strand (blue) and the RNA transcript (red).

View is from the top⁷, but rotated by 30° towards the front view. **d**, $2F_o - F_c$ electron density (blue mesh, contoured at 1.3σ) for the bridge helix (green) reveals a central, unwound region. At the bottom, a comparison of the bridge helix in Pol II (orange), aligned on the core module side, reveals an expansion by 6 Å (arrow). View from the front as in Fig. 1. **e**, The expander (raspberry) stabilizes the unwound central bridge helix (green). View as in **c**. **f**, Binding of the expander and the DNA-RNA hybrid in the active centre are mutually exclusive. Superposition of the Pol I structure with the Pol II elongation complex structure (PDB code 1Y1W) reveals clashes between the expander (raspberry) and the DNA template strand (blue) and the RNA transcript (red).

modelled contracted state of Pol I, and a C-terminal truncation of A43 that deletes the connector results in thermosensitivity *in vivo*²⁹, and is lethal when combined with other Pol I mutations³³.

Transcription regulation

Pol I dimerization inhibits assembly of the transcription initiation complex because it occludes the binding sites for the Pol I initiation

factors Rrn3 (refs 29, 34) and Rrn7 (refs 34–36) (Fig. 6c). Pol I dimers and monomers are in a concentration-dependent equilibrium *in vitro* (Extended Data Fig. 1b), but also monomers are predicted to be inactive because the expander overlaps with the binding sites for the DNA template strand and RNA (Fig. 4f). The expander binds the cleft of Pol I in solution, as shown by extensive crosslinking¹¹ (Extended Data Fig. 8d). The expander interferes with DNA loading in the active

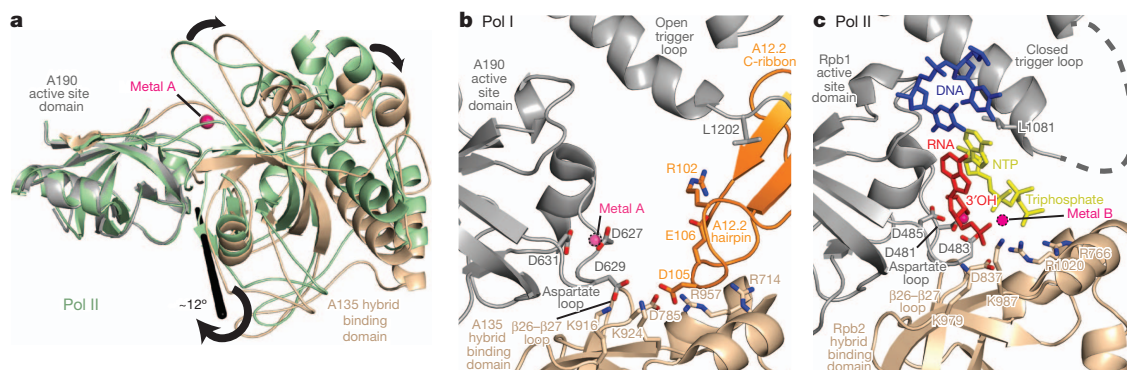


Figure 5 | Composite active site and A12.2 hairpin. **a**, Tilting between the A190 active site domain and the A135 hybrid binding domain. The active site domains of Pol I and Pol II were superposed, and the axis of rotation is indicated. The view is approximately from the side. **b**, View of the Pol I active centre from the A135 side towards the clamp. The A12.2 hairpin (orange) forms an integral part of the active site. The backbone carbonyl of Asp 627 is flipped compared to its counterpart Asp 481 in Pol II, and the side chain of

Asp 629 is rotated and interacts with Lys 916 and Lys 924 in A135. To catalyse nucleotide addition, the aspartate loop must re-establish a metal-binding conformation and move closer to the A135 loop $\beta 26$ – $\beta 27$, which may contribute to binding of metal ion B. **c**, For comparison, the active site in Pol II with DNA template (blue), RNA (red), and nucleoside triphosphate substrate analogue (yellow) and the closed trigger loop are shown²⁷.

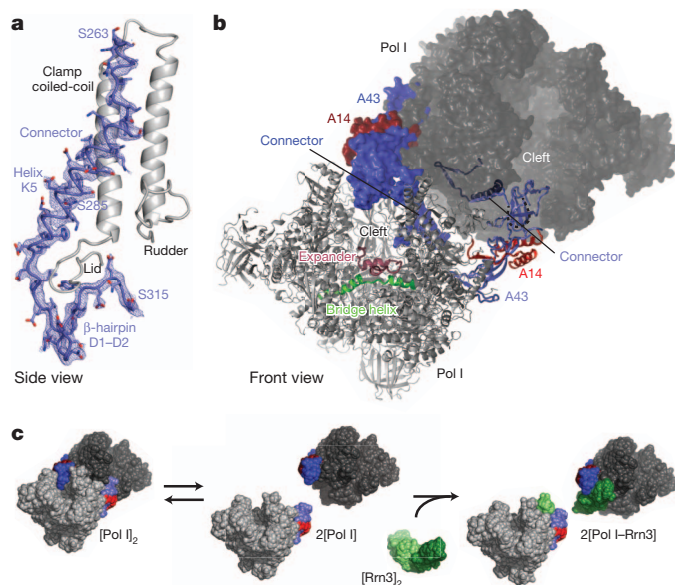


Figure 6 | Connector and Pol I dimerization. **a**, Structure of the A43 connector and corresponding $2F_o - F_c$ electron density (blue mesh, contoured at 1σ) and its binding to the coiled-coil at the inner side of the clamp (silver). **b**, Structure of the Pol I dimer and its stabilization by the A14–A43 subcomplex. One polymerase is shown as a cartoon model, the other in surface representation. The connector (blue) invades the cleft of a neighbouring polymerase. **c**, Model for Pol I initiation regulation. The binding site for the initiation factor Rrn3 (green) is occluded in the inactive Pol I dimer (silver; A14 and A43 highlighted in red and blue, respectively). Rrn3 is released from inactive Rrn3 dimers³⁴, resulting in active Pol I–Rrn3 complexes. For details compare text.

centre, because Pol I shows only weak *de novo* RNA synthesis from a tailed DNA template, in contrast to Pol II, which is much more active (Extended Data Fig. 1c). Consistent with a regulatory role of the expander, its deletion does not lead to a growth defect *in vivo* (not shown).

Pol I initiation thus requires (1) release of the connector, to generate monomers and enable Rrn3 and Rrn7 binding, (2) release of the expander, to enable DNA loading, and (3) cleft contraction, to induce a catalytically competent conformation of the active centre and enable RNA synthesis. The resulting mature DNA–RNA hybrid excludes the expander from the active centre during elongation, because Pol I is as active as Pol II when provided with a mature hybrid (Extended Data Fig. 1a).

This model for regulated Pol I initiation agrees with known functional data. First, Pol I dimers from yeast extracts are inactive in transcription^{37,38}. Second, when the connector is unavailable for Pol I dimerization, owing to a fusion with Rrn3, Pol I regulation is impaired *in vivo*³. Third, A43 phosphorylation apparently counteracts Pol I dimerization and promotes Rrn3 binding, because (1) A43 binds Rrn3 (ref. 29) and is phosphorylated in Pol I–Rrn3 complexes³⁹, (2) A43 phosphorylation sites⁴⁰ at Ser 208, Ser 220, Ser 262/263 and Ser 285 are exposed in our structure and face Rrn3 (ref. 34), and (3) Pol I dephosphorylation prevents Rrn3 binding and results in low transcription activity³⁹, and extracts from non-growing cells lack active Pol I–Rrn3 complexes³⁸. The same regulatory mechanism apparently exists in multicellular eukaryotes, because the expander is well conserved, and the connector is partially conserved (Extended Data Fig. 8b, c). Consistent with this, inactive human Pol I may be dimeric, whereas human Pol I complexes containing the Rrn3 homologue transcription initiation factor-1A (TIF-1A) are active⁴¹. Like Rrn3, human TIF-1A binds A43 (ref. 42), which may be dephosphorylated during the cell cycle by the Cdc14 phosphatase that inhibits Pol I (ref. 43).

Conclusions

The Pol I structure enables a detailed structure–function analysis of rRNA production, the first step in ribosome biogenesis. It also elucidates the evolution of related cellular RNA polymerases. Pol I and III apparently acquired a composite active site for intrinsic RNA proof-reading, 3′ processing, and chain termination, to enable accurate, high-level production of stable RNA products. The composite active site probably impairs extensive RNA backtracking because the binding site for backtracked RNA⁴⁴ is occupied by A12.2. Further backtracking may occur during termination and enable cleft expansion and hybrid release, as suggested for the bacterial enzyme⁴⁵ and for Pol III (ref. 46). By contrast, the tunable active site of Pol II allows for accommodation of backtracked RNA in the pore, Pol II-specific regulation during the elongation phase, and 3′ RNA processing on the polymerase surface by the machinery for pre-mRNA cleavage and polyadenylation.

The most intriguing finding from this work, however, relates to a possible general mechanism of transcription regulation. We observe an expanded, inhibited conformation of a eukaryotic RNA polymerase that resembles a conformation of bacterial RNA polymerase associated with inactive states^{25,26}. Because the polymerase core is highly conserved, all cellular RNA polymerases can probably inter-convert between an expanded inactive state and a contracted active state. State conversion will change the catalytic properties of the enzyme, because the active site is formed at the interface between the mobile core and shelf modules. Only the contracted state contains the active site in a catalytically competent conformation and binds the hybrid tightly. The conversion between polymerase states may be regulated allosterically by factors that bind at the interface of the core and shelf modules. Factors may bind in the pore, as observed for TFIIS¹⁴ and Gfh1 (ref. 25), or near the RNA exit channel, as observed for Rrn3 (refs 29, 34), the Pol II initiation factor TFIIB that stimulates RNA chain initiation allosterically⁴⁷, and the Pol II coactivator Mediator⁴⁸. The bacterial regulator catabolite activator protein⁴⁹ and the growth regulator ppGpp also bind at the RNA exit channel, and the latter was suggested to influence polymerase activity by modulating the core–shelf interface⁵⁰.

METHODS SUMMARY

A *S. cerevisiae* strain with a C-terminal affinity-purification tag on A190 was grown in a 200l fermenter. Endogenous Pol I was prepared as described⁹ with several modifications (Methods). Crystals were grown by vapour diffusion at 20 °C using 10% PEG 4000, 300 mM ammonium acetate, 50 mM HEPES, pH 7.5, and 5 mM TCEP as precipitant, and improved by microseeding in 8% PEG 4000, 300 mM ammonium acetate, 50 mM imidazole, pH 7.0, and 5 mM TCEP. Diffraction data were collected under cryo-conditions at Swiss Light Source beamlines PX1 and PX2, and at European Synchrotron Radiation Facility beamlines ID23-1 and ID29 using Pilatus 6M detectors. The structure was solved by multiwavelength anomalous diffraction from 14 intrinsic zinc ions in the Pol I dimer. Solvent flattening, non-crystallographic symmetry averaging, and phase extension to 2.8 Å resolution led to a detailed electron density map that allowed building and refinement of an atomic model.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 June; accepted 24 September 2013.

Published online 23 October 2013.

1. Roeder, R. G. & Rutter, W. J. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234–237 (1969).
2. Miller, O. L. Jr & Beatty, B. R. Visualization of nucleolar genes. *Science* **164**, 955–957 (1969).
3. Laferté, A. *et al.* The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes Dev.* **20**, 2030–2040 (2006).
4. Drygin, D., Rice, W. G. & Grummt, I. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu. Rev. Pharmacol. Toxicol.* **50**, 131–156 (2010).

5. Schultz, P., Celia, H., Riva, M., Sentenac, A. & Oudet, P. Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. *EMBO J.* **12**, 2601–2607 (1993).
6. Bischler, N. *et al.* Localization of the yeast RNA polymerase I-specific subunits. *EMBO J.* **21**, 4136–4144 (2002).
7. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001).
8. Cheung, A. C. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* **149**, 1431–1437 (2012).
9. Kuhn, C. D. *et al.* Functional architecture of RNA polymerase I. *Cell* **131**, 1260–1272 (2007).
10. Geiger, S. R. *et al.* RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol. Cell* **39**, 583–594 (2010).
11. Jennebach, S., Herzog, F., Aebbersold, R. & Cramer, P. Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res.* **40**, 5591–5601 (2012).
12. Korkhin, Y. *et al.* Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* **7**, e1000102 (2009).
13. Hirata, A. *et al.* Archaeal RNA polymerase subunits E and F are not required for transcription *in vitro*, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol. Microbiol.* **70**, 623–633 (2008).
14. Kettenberger, H., Armache, K. J. & Cramer, P. Architecture of the RNA polymerase II–TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).
15. Buhler, J. M., Sentenac, A. & Fromageot, P. Isolation, structure, and general properties of yeast ribonucleic acid polymerase A (or I). *J. Biol. Chem.* **249**, 5963–5970 (1974).
16. Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D. & Cramer, P. Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J. Biol. Chem.* **286**, 18701–18707 (2011).
17. Alic, N. *et al.* Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. *Proc. Natl Acad. Sci. USA* **104**, 10400–10405 (2007).
18. Prescott, E. M. *et al.* Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc. Natl Acad. Sci. USA* **101**, 6068–6073 (2004).
19. Arimbasseri, A. G., Rijal, K. & Marais, R. J. Transcription termination by the eukaryotic RNA polymerase III. *Biochim. Biophys. Acta* **1829**, 318–330 (2013).
20. Van Mullem, V., Landrieux, E., Vandenhaute, J. & Thuriaux, P. Rpa12p, a conserved RNA polymerase I subunit with two functional domains. *Mol. Microbiol.* **43**, 1105–1113 (2002).
21. Beckouët, F. *et al.* Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol. Cell Biol.* **28**, 1596–1605 (2008).
22. Chen, Z. A. *et al.* Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
23. Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
24. Fernández-Torero, C. *et al.* Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J.* **29**, 3762–3772 (2010).
25. Tagami, S. *et al.* Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468**, 978–982 (2010).
26. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431–441 (2013).
27. Cheung, A. C., Sainsbury, S. & Cramer, P. Structural basis of initial RNA polymerase II transcription. *EMBO J.* **30**, 4755–4763 (2011).
28. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954 (2006).
29. Peyroche, G. *et al.* The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J.* **19**, 5473–5482 (2000).
30. Armache, K. J., Mitterweger, S., Meinhardt, A. & Cramer, P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–7134 (2005).
31. Jasiak, A. J., Armache, K. J., Martens, B., Jansen, R. P. & Cramer, P. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol. Cell* **23**, 71–81 (2006).
32. Smid, A., Riva, M., Bouet, F., Sentenac, A. & Carles, C. The association of three subunits with yeast RNA polymerase is stabilized by A14. *J. Biol. Chem.* **270**, 13534–13540 (1995).
33. Beckouët, F., Mariotte-Labarre, S., Peyroche, G., Nogi, Y. & Thuriaux, P. Rpa43 and its partners in the yeast RNA polymerase I transcription complex. *FEBS Lett.* **585**, 3355–3359 (2011).
34. Blattner, C. *et al.* Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev.* **25**, 2093–2105 (2011).
35. Knutson, B. A. & Hahn, S. Yeast Rrn7 and human TAF1B are TFIIB-related RNA polymerase I general transcription factors. *Science* **333**, 1637–1640 (2011).
36. Naidu, S., Friedrich, J. K., Russell, J. & Zomerdijs, J. C. TAF1B is a TFIIB-like component of the basal transcription machinery for RNA polymerase I. *Science* **333**, 1640–1642 (2011).
37. Milkereit, P., Schultz, P. & Tschochner, H. Resolution of RNA polymerase I into dimers and monomers and their function in transcription. *Biol. Chem.* **378**, 1433–1443 (1997).
38. Milkereit, P. & Tschochner, H. A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *EMBO J.* **17**, 3692–3703 (1998).
39. Fath, S. *et al.* Differential roles of phosphorylation in the formation of transcriptional active RNA polymerase I. *Proc. Natl Acad. Sci. USA* **98**, 14334–14339 (2001).
40. Gerber, J. *et al.* Site specific phosphorylation of yeast RNA polymerase I. *Nucleic Acids Res.* **36**, 793–802 (2008).
41. Miller, G. *et al.* hRRN3 is essential in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters. *EMBO J.* **20**, 1373–1382 (2001).
42. Yuan, X., Zhao, J., Zentgraf, H., Hoffmann-Rohrer, U. & Grummt, I. Multiple interactions between RNA polymerase I, TIF-1A and TAF_I subunits regulate preinitiation complex assembly at the ribosomal gene promoter. *EMBO Rep.* **3**, 1082–1087 (2002).
43. Clemente-Blanco, A. *et al.* Cdc14 inhibits transcription by RNA polymerase I during anaphase. *Nature* **458**, 219–222 (2009).
44. Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).
45. Epstein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S. & Nudler, E. An allosteric path to transcription termination. *Mol. Cell* **28**, 991–1001 (2007).
46. Nielsen, S., Yuzenkova, Y. & Zenkin, N. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* **340**, 1577–1580 (2013).
47. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. *Nature* **493**, 437–440 (2013).
48. Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C. & Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription *in vivo*. *Science* **331**, 1451–1454 (2011).
49. Lawson, C. L. *et al.* Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* **14**, 10–20 (2004).
50. Zuo, Y., Wang, Y. & Steitz, T. A. The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol. Cell* **50**, 430–436 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Kuhn, who obtained the first Pol I crystals. We thank C. Bäjén, S. Benkert, S. Geiger, S. Jennebach, T. Gubbey and K. Maier. We thank the crystallization facility (Conti Department) and the Jentsch Department of the Max-Planck-Institut für Biochemie. Part of this work was performed at the European Synchrotron Radiation Facility at Grenoble, France, and at the Swiss Light Source at the Paul-Scherrer-Institut, Villigen, Switzerland. C.E. was supported by a PhD student fellowship of the Boehringer Ingelheim Fonds, the Elite Network Bavaria program ‘Protein Dynamics in Health and Disease’, and the Graduate Research Academy ‘RNA Biology’ of SFB960. S.S. was supported by a postdoctoral fellowship of the Alexander-von-Humboldt Foundation. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB646, TR5, GraKo1721, SFB960, CIPSM, NIM), an Advanced Grant of the European Research Council, the Jung-Stiftung, and the Vallee Foundation.

Author Contributions C.E. planned and carried out experiments and crystal structure determination. S.S. advised on experimental and crystallographic work. A.C.C. performed computational crystallographic analysis. D.K. contributed to computational crystallography and model building. P.C. designed and supervised research and prepared the manuscript, with contributions from all authors.

Author Information Atomic coordinates and structure factors of the Pol I crystal structure have been deposited with the Protein Data Bank under accession number 4C2M. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (cramer@LMB.uni-muenchen.de).

METHODS

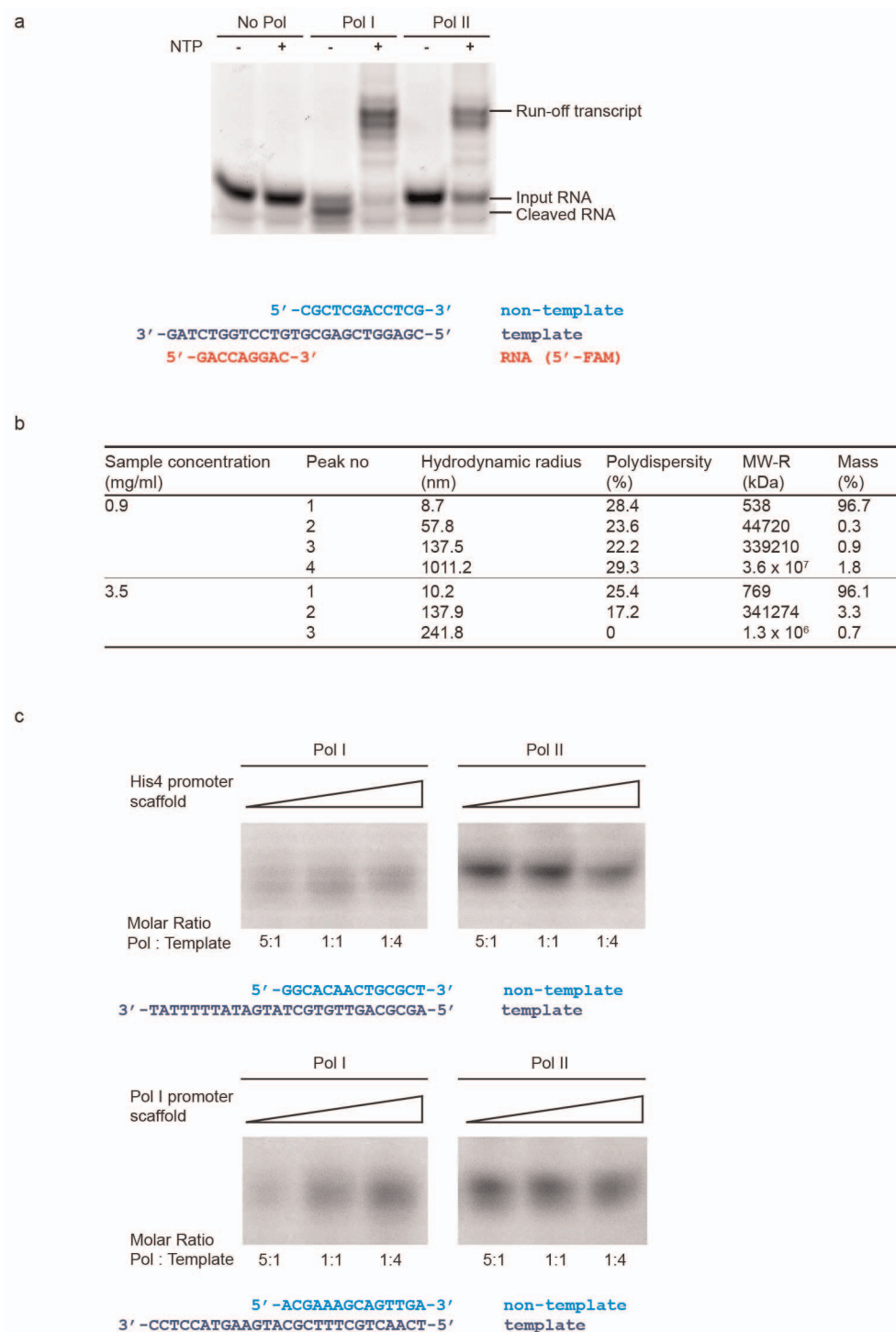
Pol I preparation. Pol I enzyme used for crystallization was essentially obtained as described⁹, using a protocol that was based on a procedure originally established by the Tschochner laboratory⁵¹, but with several changes. Pol I was purified from the expression-optimized strain CB010 expressing a C-terminal Flag/10×histidine-tagged subunit A190. Cells were grown to an attenuation (*D*) at 600 nm of 2.0 in YPD medium in a 200 l fermenter and collected using a continuous-flow centrifuge. Cells were suspended and flash-frozen in buffer A (150 mM HEPES, pH 7.8, 60 mM MgCl₂, 20% (v/v) glycerol, 5 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM benzamidine, 60 μM leupeptine and 200 μM pepstatine). For purification, 500 g of yeast cells were lysed by bead beating in buffer A after the addition of ammonium sulphate to 400 mM at 4 °C. The lysate was cleared by centrifugation (30 min at 10,000g) and ultracentrifugation (90 min at 30,000g) and dialysed overnight against buffer B (50 mM potassium acetate, 20 mM HEPES, pH 7.8, 1 mM MgCl₂, 10% (v/v) glycerol, 10 mM β-mercaptoethanol, 1 mM PMSF and 1 mM benzamidine). After centrifugation (60 min at 18,500g) the pellet was re-suspended in buffer C (1.5 M potassium acetate, 20 mM HEPES, pH 7.8, 1 mM MgCl₂, 10 mM imidazole, 10% (v/v) glycerol, 10 mM β-mercaptoethanol, 1 mM PMSF and 1 mM benzamidine) and incubated with 8 ml Ni-NTA beads (Qiagen) for 4 h. Bound protein was washed with five column volumes (CV) buffer C, followed by 5 CV buffer D (300 mM potassium acetate, 20 mM HEPES, pH 7.8, 1 mM MgCl₂, 25 mM imidazole, 10% (v/v) glycerol, 10 mM β-mercaptoethanol) and eluted in 5 CV buffer E (buffer D but containing 200 mM imidazole). The eluate was loaded on a 10/100 MonoQ anion exchange column (GE Healthcare) and eluted with a gradient from 0.3 to 2.0 M potassium acetate in buffer F (20 mM HEPES, pH 7.8, 1 mM MgCl₂, 10% (v/v) glycerol and 5 mM DTT). Eluted Pol I was diluted from 1.1 to 0.2 M potassium acetate, loaded on a 5/50 MonoS cation exchange column (GE Healthcare) and eluted with a gradient from 0.2 to 1.0 M potassium acetate in buffer F including a plateau at 350 mM potassium acetate. Pol I eluted at 490 mM, was concentrated to 500 μl and applied to a Superose 6 10/300 size exclusion column (GE Healthcare) in buffer G (60 mM ammonium sulphate, 5 mM HEPES, pH 7.8, 1 mM MgCl₂, 10 μM ZnCl₂, 5 mM DTT). A single peak contained up to 1.8 mg of homogenous and pure Pol I.

Pol I crystallization. Freshly purified Pol I was subjected to commercial screens and to crystallization conditions that had previously produced RNA polymerase crystals from different organisms. MRC-2 microplates (96-well format; SWISSCI) were set up with 100 nl sitting drops using the Phoenix pipetting device (Art Robbins). Initial crystals were obtained in 10% PEG 4000, 300 mM ammonium acetate, 50 mM HEPES, pH 7.5, and 5 mM TCEP at 20 °C. These crystals formed thin plates that poorly diffracted X-rays. Optimization and micro-seeding yielded brick-like crystals with a size of up to 300 × 60 × 40 μm using as reservoir solution 8% PEG 4000, 300 mM ammonium acetate, 50 mM imidazole, pH 7.0, and 5 mM TCEP. After a two-step transfer to the same buffer containing an additional 20% PEG400, crystals were flash-frozen and stored in liquid nitrogen.

Pol I crystal structure determination. Diffraction data were collected at the Swiss Light Source in Villigen, Switzerland, on beamlines PX1 and PX2 as well as at the European Synchrotron Radiation Facility in Grenoble, France, on beamlines ID23-1 and ID29 using Pilatus 6M detectors. Data were processed with XDS⁵² and showed P1 symmetry (Extended Data Table 2a, b). A two-fold non-crystallographic symmetry (NCS) axis was detected by calculation of a self-rotation function. Molecular

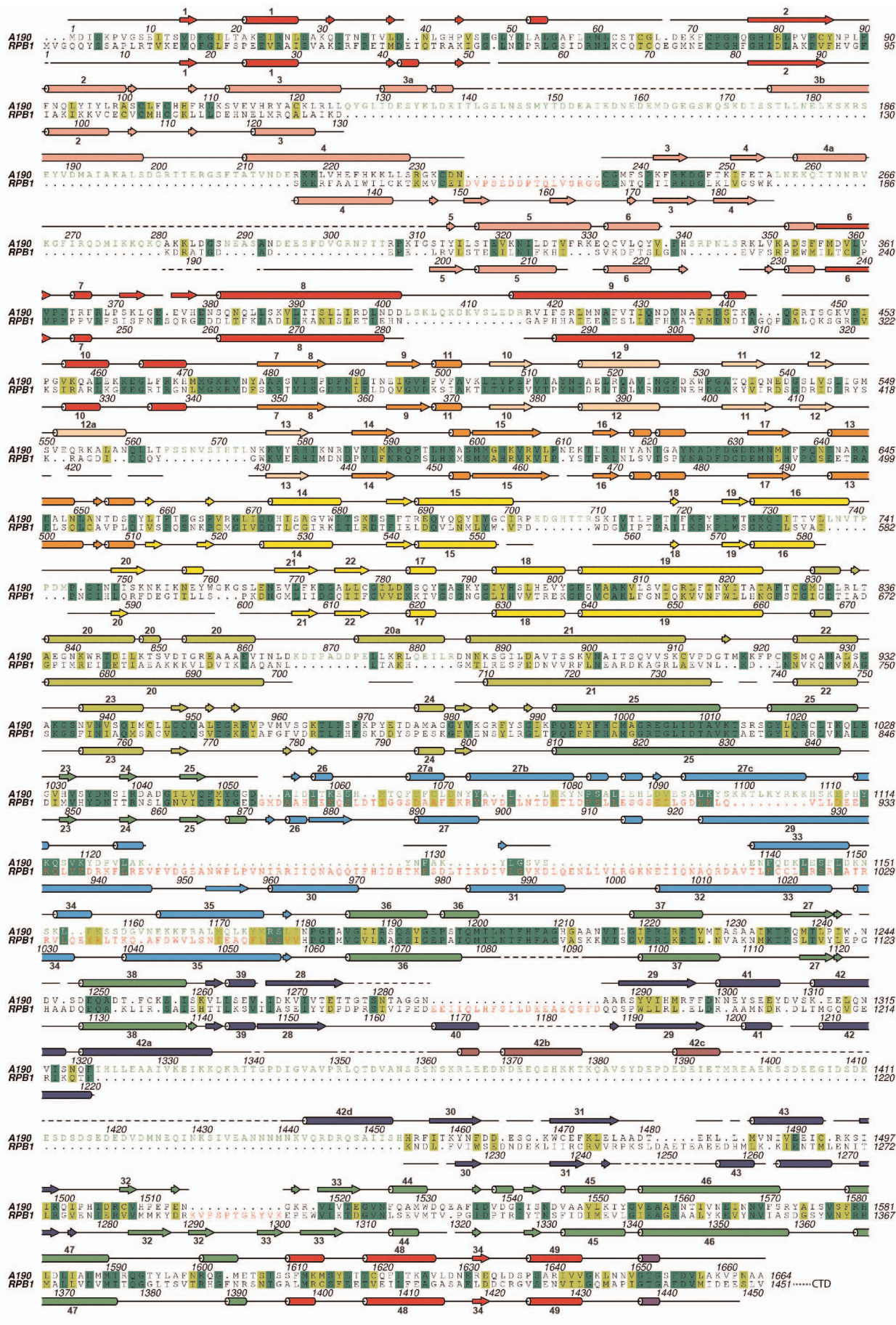
replacement using as search models Pol II, a Pol I homology model¹¹ or a Pol I electron microscopic envelope⁹ failed. We could, however, use anomalous diffraction from intrinsically bound zinc ions for phasing. After performing a fluorescence scan, three data sets were collected from a single crystal at the peak, inflection, and low-energy remote wavelengths (Extended Data Table 2b). From the peak data set, the positions of 14 zinc ions (7 in each Pol I complex) were determined with SHELX C and D⁵³. The occupancies calculated by SHELX were between 100% and 58% for the first 14 sites, followed by a drop to ~25% for additional peaks that were not considered. Zinc positions were refined using autoSHARP⁵⁴ in a multiwavelength anomalous diffraction experiment using four wavelengths (Extended Data Table 2a, b). Two Pol II-based homology models of the Pol I core were fitted to match the experimental zinc positions using the program MOLOC. Initial phases at 3.5 Å resolution were obtained with autoSHARP followed by solvent flattening with SOLOMON⁵⁴ using the positioned homology models as an initial mask. The program PARROT⁵⁵ performed automated NCS averaging and extended phases to 2.8 Å resolution. An atomic model of Pol I was built with COOT⁵⁶ starting from the positioned homology models. The PHENIX suite⁵⁷ was used to refine the structure with manually assigned rigid body groups considering domain boundaries, and independent TLS parameterization for a total of 24 TLS groups. NCS restraints with groups defined according to the rigid bodies and individual atomic displacement parameters were applied during refinement. In the later stages of refinement, prominent water molecules and two bound sulphate ions were placed and validated by visual inspection. The final model has an *R*_{free} factor of 21.0% at 2.8 Å resolution and excellent stereochemistry (Extended Data Table 2a). For visualization, secondary structure was assigned with DSSP⁵⁸ and by inspection.

51. Tschochner, H. A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. *Proc. Natl Acad. Sci. USA* **93**, 12914–12919 (1996).
52. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
53. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
54. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
55. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66**, 470–478 (2010).
56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
57. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
58. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
59. Bischof, N. *et al.* Specific interaction and two-dimensional crystallization of histidine tagged yeast RNA polymerase I on nickel-chelating lipids. *Biophys. J.* **74**, 1522–1532 (1998).
60. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
61. Geiger, S. R., Kuhn, C. D., Leidig, C., Renkawitz, J. & Cramer, P. Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions. *Acta Crystallogr. F* **64**, 413–418 (2008).
62. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
63. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).



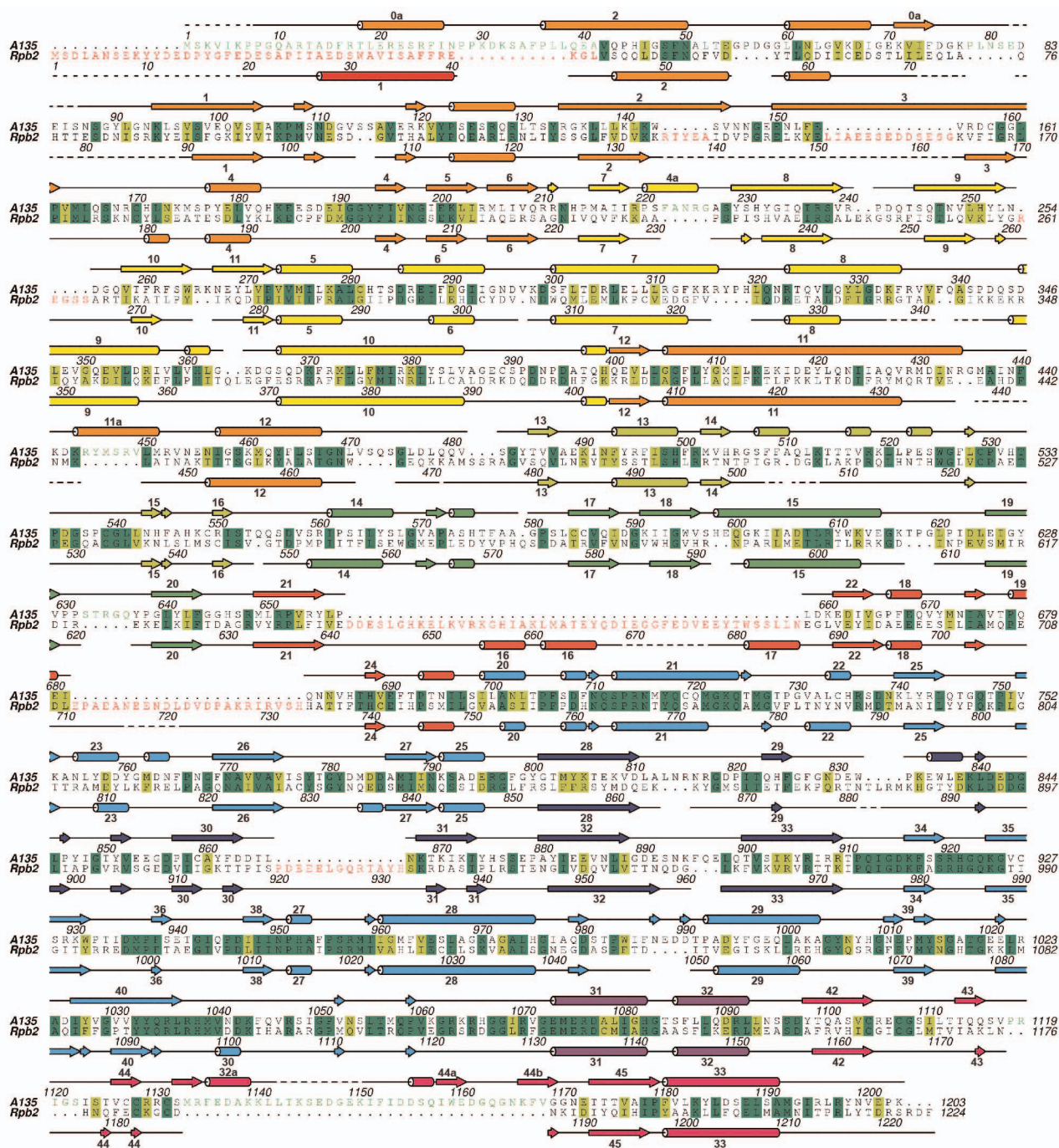
Extended Data Figure 1 | Activity of purified Pol I *in vitro*. **a**, Pol I is active in RNA extension and cleavage. Purified *S. cerevisiae* Pol I and Pol II can extend RNA in a DNA–RNA scaffold in the presence of nucleoside triphosphate substrates (NTPs). In the absence of nucleoside triphosphate substrates, Pol I cleaves RNA. Transcription assays were performed as described⁹ using the DNA–RNA scaffold shown. After running a 20% acrylamide UREA gel, RNA was detected by fluorescence. **b**, Dynamic light scattering is consistent with a concentration-dependent Pol I dimer–monomer equilibrium in solution. Note that the sample with the higher Pol I concentration shows an increased

hydrodynamic radius and apparent (estimated) molecular mass for the predominant peak by mass that accounts for over 96%. The method cannot distinguish monomeric and dimeric species. Thus, the estimated molecular mass observed at 3.5 mg ml^{-1} Pol I concentration (769 kDa) arises from a mixture of monomers (590 kDa) and dimers (1,180 kDa). **c**, *De novo* RNA synthesis activity on tailed DNA template. Assays were performed as described⁴⁷ using the DNA template shown. After running a 20% acrylamide UREA gel, *de novo* synthesized, radioactive RNA was detected by phosphorimaging.



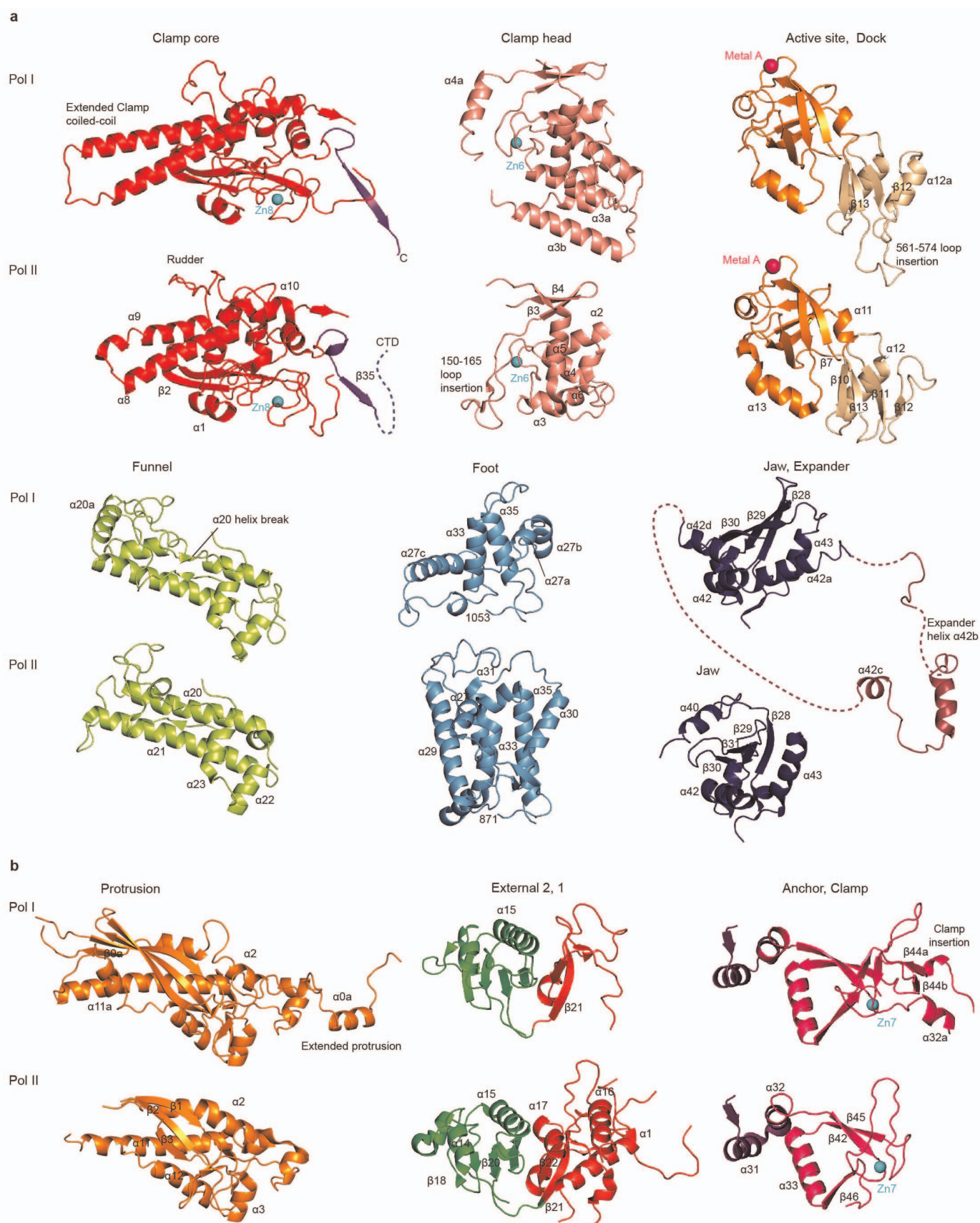
Extended Data Figure 2 | Structure-based alignment of A190 and Rpb1.
Invariant and conserved residues are highlighted in green and light green, respectively. Secondary structure elements are indicated above and below the

alignment for A190 and Rpb1, respectively (cylinders, helices; arrows, strands). Residues that form different folds in Pol I or form Pol-II-specific folds are in green or red, respectively (compare Fig. 1c).



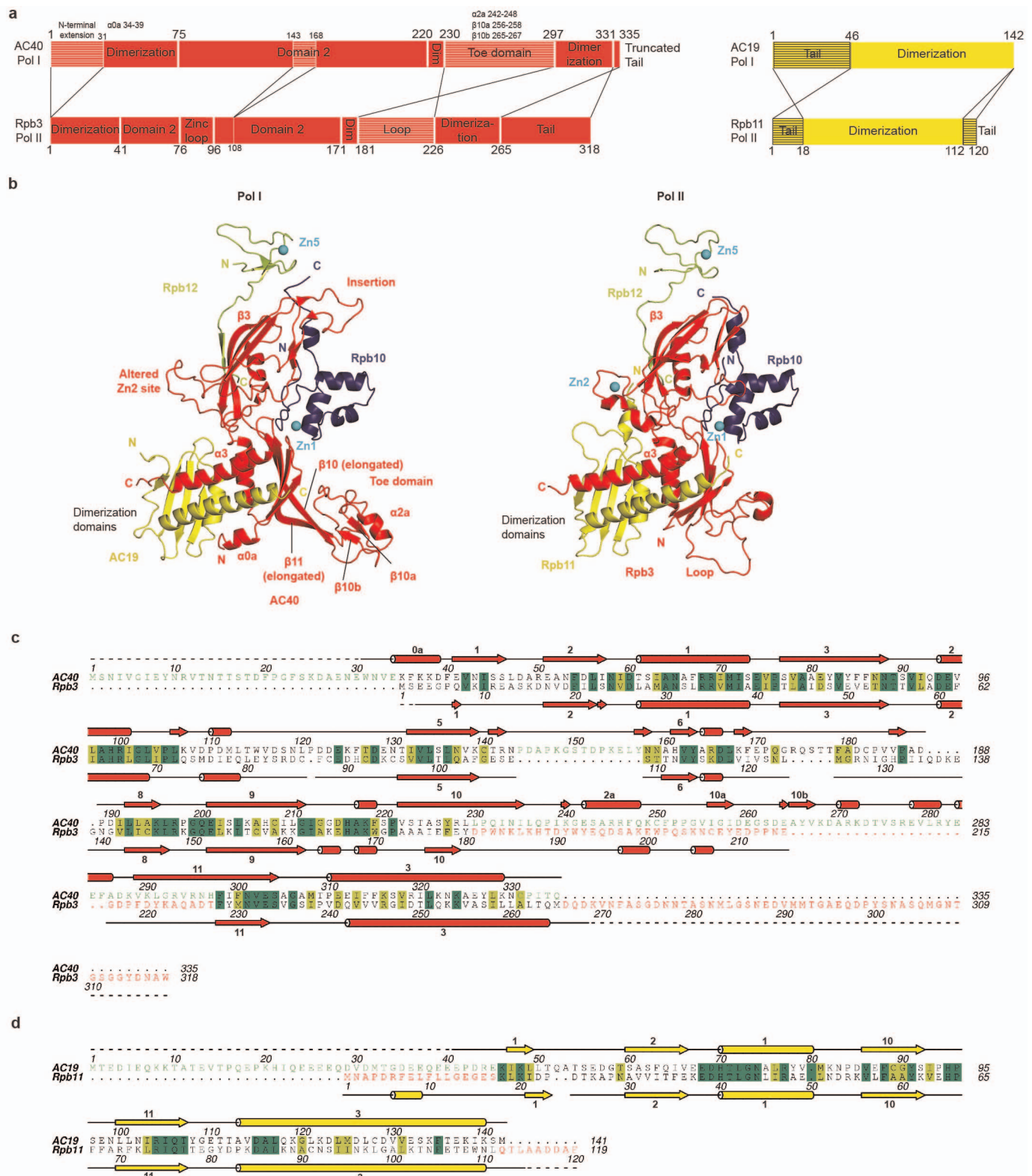
Extended Data Figure 3 | Structure-based alignment of A135 and Rpb2. Invariant and conserved residues are highlighted in green and light green, respectively. Secondary structure elements are indicated above and below the

alignment for A135 and Rpb2, respectively (cylinders, helices; arrows, strands). Residues that form different folds in Pol I or form Pol-II-specific folds are in green or red, respectively (compare Fig. 1c).



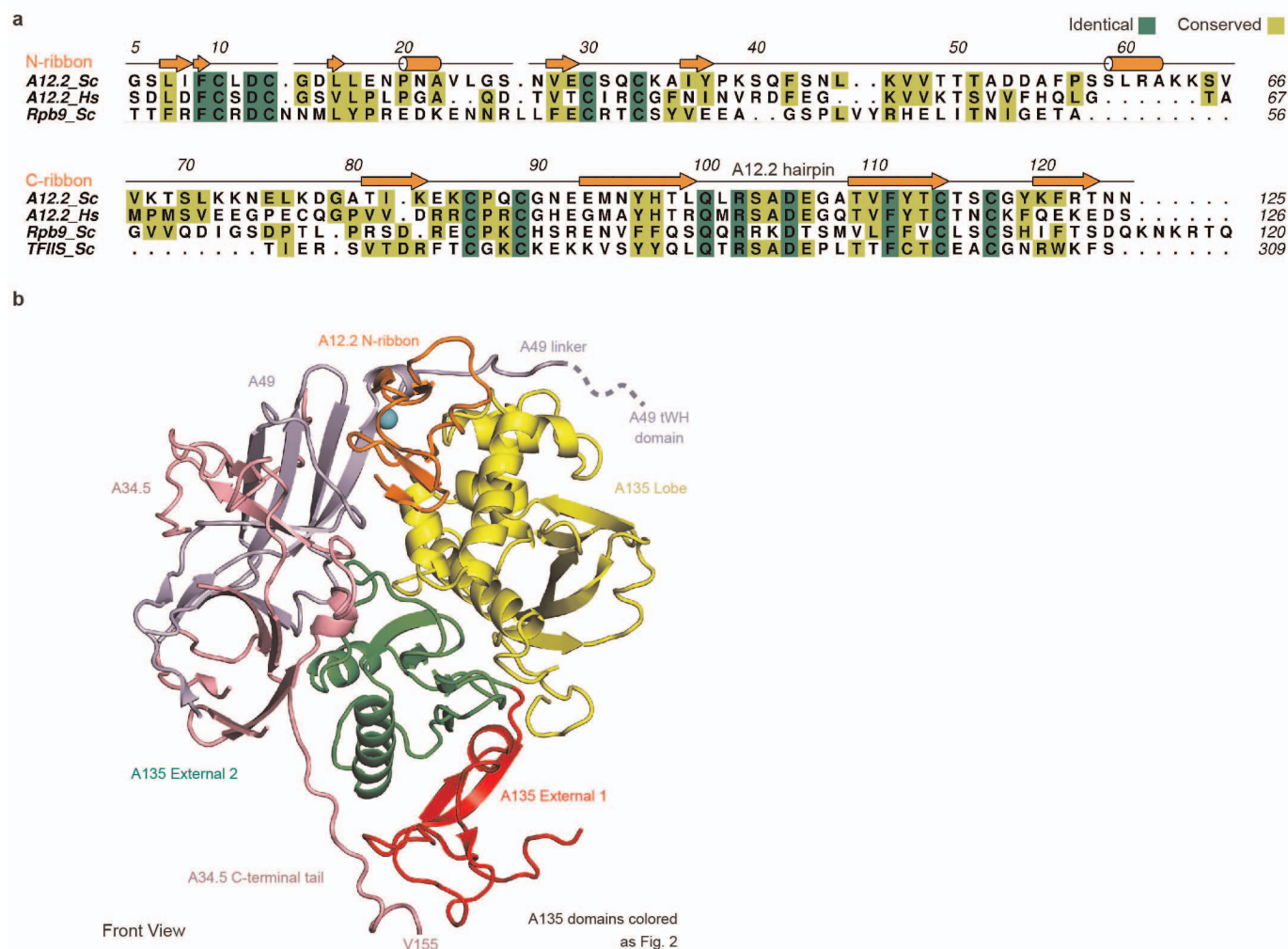
Extended Data Figure 4 | Detailed comparison of A190–A135 with Rpb1–Rpb2. **a**, Comparison of A190 domain structures (top) that differ significantly from their corresponding Pol II regions (bottom). Labelling of corresponding secondary structure elements is as for Pol II (ref. 7). New or lacking secondary

structure elements are labelled. New elements were named according to the preceding Pol II element with small letters added alphabetically for subsequent elements. **b**, Comparison of A135 domain structures (top) that differ significantly from their corresponding Pol II regions (bottom) as in Extended Data Fig. 3.



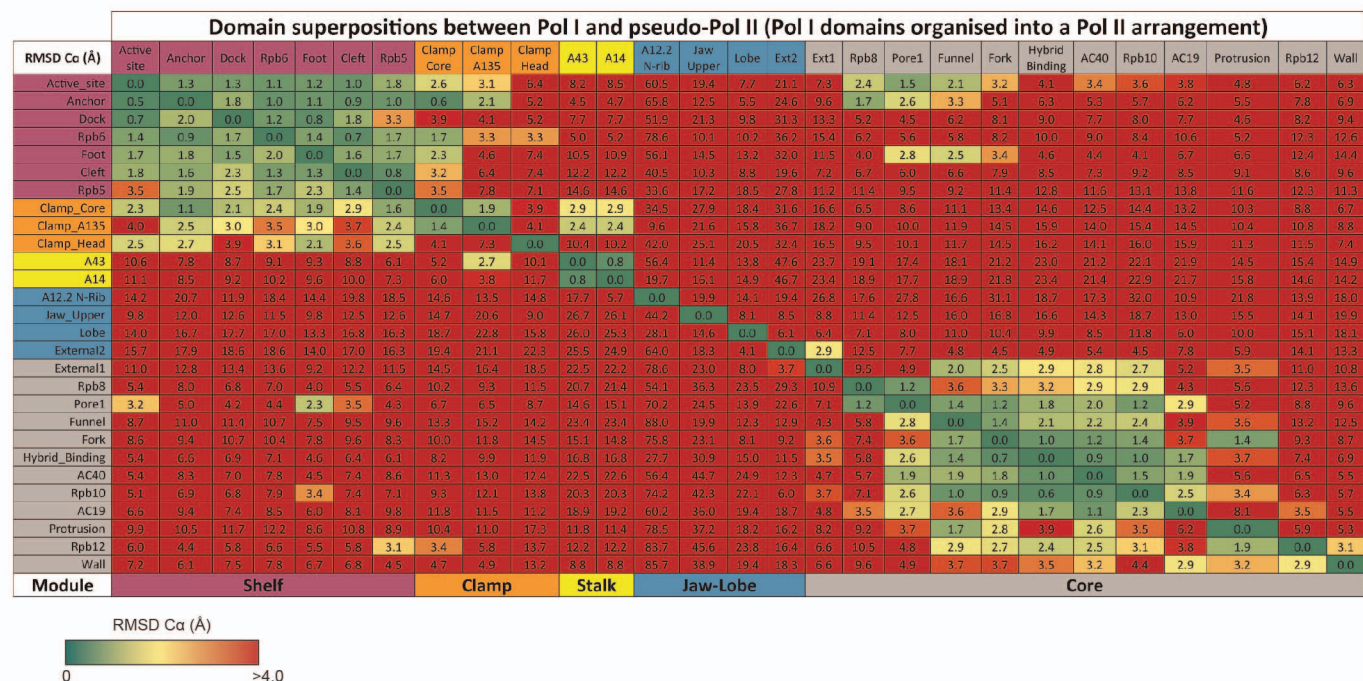
Extended Data Figure 5 | Structure of the subassembly AC40–AC19–Rpb10–Rpb12. **a**, Schematic of domain structures. AC40 and AC19 were aligned with homologous Pol II subunits Rpb3 and Rpb11, respectively. Organization and annotation as in Fig. 2a. **b**, Ribbon view from the ‘back’ of the enzyme⁷ (left, Pol I; right, Pol II). **c**, **d**, Structure-based alignments of AC40 and

Rpb3 (**c**) and of AC19 and Rpb11 (**d**). Invariant and conserved residues are highlighted in green and light green, respectively. Secondary structure elements are indicated above and below the alignment for AC40 and Rpb3, respectively (cylinders, helices; arrows, strands). Residues that form different folds in Pol I or form Pol-II-specific folds are in green or red, respectively (compare Fig. 1c).



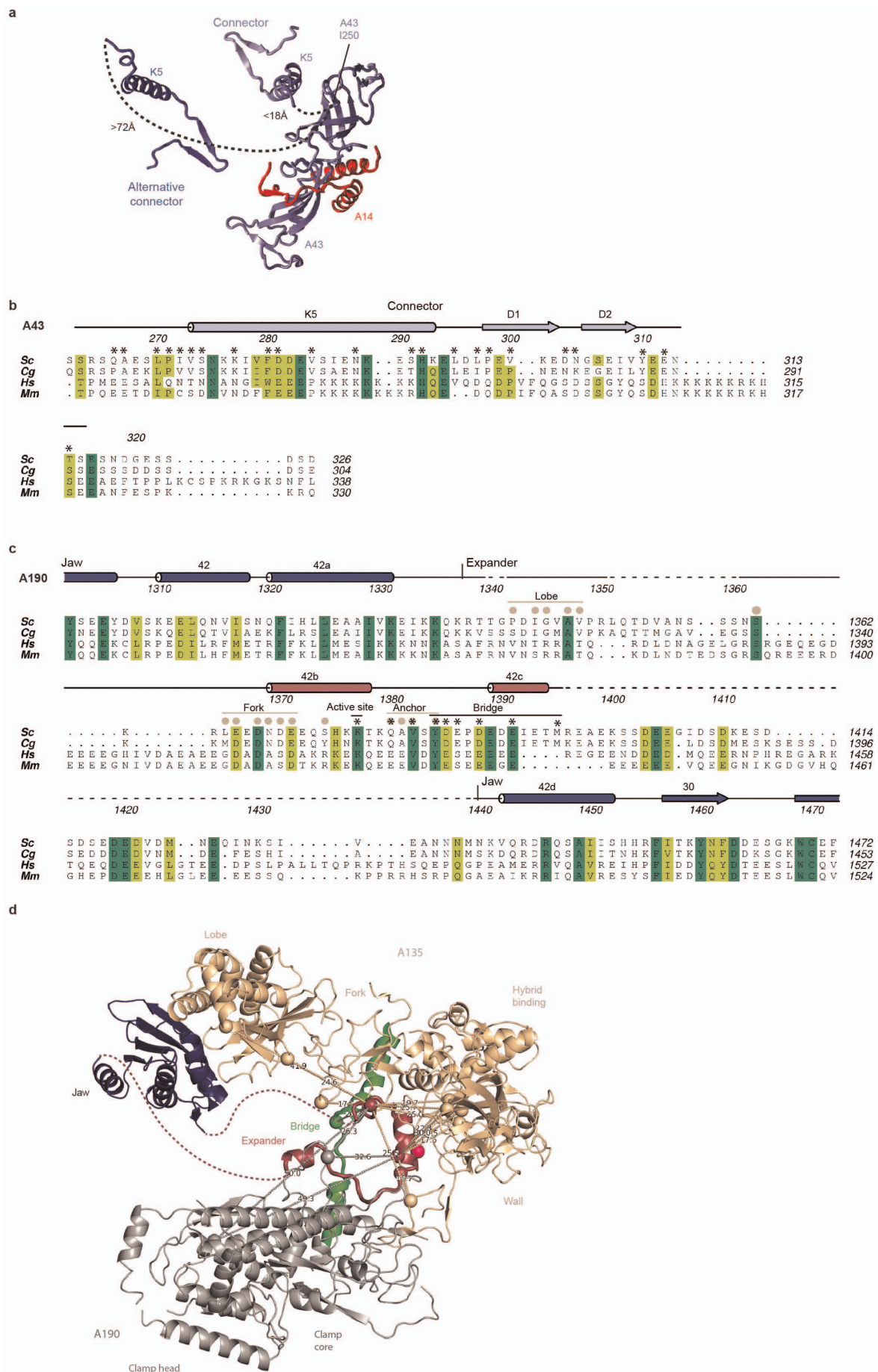
Extended Data Figure 6 | A12.2 and A49–A34.5. **a**, Subunit A12.2 amino acid sequence alignment. The sequence of yeast *S. cerevisiae* A12.2 (Sc) was aligned with that of the human subunit (*Homo sapiens*; Hs), and the N- and C-terminal domains were aligned with their counterparts in the Pol II subunit

Rpb9 and the Pol II elongation factor TFIIS¹⁴, respectively. **b**, Interaction of A49–A34.5 with Pol I core domains. The view is from the front of the enzyme⁷ (Fig. 1). Different Pol I subunit domains are coloured as in Figs 2 and 3.



Extended Data Figure 7 | Shift in domain positions between Pol I and Pol II. Individual domain fragments of Pol I were isolated and superposed onto their counterparts in Pol II and subsequently recombined to form a 'pseudo-Pol II' model. This pseudo-Pol II model was then superposed onto the complete Pol I structure over a single common domain (indicated by column headings).

The resulting root mean squared deviation (r.m.s.d.) Cα (Å) value for every individual domain (indicated by row headings) was calculated using PyMol and coloured according to value (green to yellow = 0.0 to 3.0 Å, orange = 3.0 to 4.0 Å, and red >4.0 Å).



Extended Data Figure 8 | Structure and conservation of the connector and expander. **a**, The connector binds the adjacent polymerase. The electron density is not continuous between the end of the A43 N-terminal part (residue Ile 250) and the beginning of the A43 connector in the adjacent polymerase. The connector cannot be assigned to A43 from the same polymerase because the observed distance between the last residue in the stalk and the first residue in the connector (>72 Å) cannot be spanned with only 12 residues. By contrast, the distance from Ile 250 to the connector in the cleft of the neighbouring polymerase is 18 Å and easily spanned by the missing residues. We note that dimeric forms of Pol I were previously observed in two-dimensional arrays⁵⁹. **b**, Amino acid sequence alignment of the connector region in *S. cerevisiae* (Sc); *Candida glabrata* (Cg); *Homo sapiens* (Hs) and *Mus musculus* (Mm). Secondary structure elements are indicated above the alignment (K5 helix, cylinder;

strands D1, D2; arrows). Residues that are involved in the interface with the Pol I clamp and cleft and showed a buried surface area in excess of 40 Å^2 are indicated with an asterisk above the alignment. Buried surfaces were calculated with the PISA server⁶⁰. A structure-based alignment of the A14–A43 stalk residues was published^{9,61} and is not included here. **c**, Amino acid sequence alignment of the expander region in *S. cerevisiae*, *C. glabrata*, *H. sapiens* and *M. musculus*. Secondary structure elements are indicated above the alignment (helices, cylinders; strands, arrows). Residues that are involved in the interface with the Pol I cleft and showed a buried surface area in excess of 40 Å^2 are indicated with an asterisk above the alignment. Buried surfaces were calculated with the PISA server⁶⁰. **d**, Previously obtained crosslinks¹¹ of the expander element map to the Pol I cleft. Crosslinks to A190 and A135 are indicated in grey and wheat respectively.

Extended Data Table 1 | RNA polymerase subunits in *S. cerevisiae*

RNA polymerase	Pol I	Pol II	Pol III [*]
Core enzyme subunits	A190	Rpb1 (22.2%) [†]	C160
	A135	Rpb2 (25.9%) [†]	C128
	AC40	Rpb3 (18.9%) [†]	AC40
	AC19	Rpb11 (20.0%) [†]	AC19
	A12.2 [‡]	Rpb9 [‡] (18.3%) [†]	C11 [‡]
	Rpb5	Rpb5	Rpb5
	Rpb6	Rpb6	Rpb6
	Rpb8	Rpb8	Rpb8
	Rpb10	Rpb10	Rpb10
	Rpb12	Rpb12	Rpb12
Stalk subcomplex A14/A43	A14	Rpb4 (1.8%) [†]	C17
	A43	Rpb7 (12.3%) [†]	C25
TFIIF-like subcomplex	A49	Tfg1 (TFIIF α)	C37
	A34.5	Tfg2 (TFIIF β)	C53

* Pol III also contains the trimeric subcomplex C82–C34–C31 that was not included here.

[†] Percentage amino acid residues of the Pol II subunit that are identical in the corresponding Pol I subunit, based on the structure-based alignments in Extended Data Figs 2, 3 and 5 and published data⁹.

[‡] The C-terminal zinc-ribbon domain of A12.2 and C11 corresponds to domain III of the Pol II elongation and transcript cleavage stimulatory factor TFIIS.

Extended Data Table 2 | Diffraction data and statistics

a		Crystal I-IV			
Data collection*					
Space group		P1			
Cell dimensions					
a, b, c (Å)		122.7, 139.0, 209.6			
α, β, γ (°)		108.1, 95.4, 93.8			
Resolution (Å)		40-2.8 (2.87-2.80) [†]			
R_{sym}		24.0 (201.8)			
$I/\sigma I$		9.61 (1.52)			
Completeness (%)		99.9 (99.8)			
Redundancy		13.1 (11.5)			
$CC_{(1/2)}^{\ddagger}$ (%)		99.6 (54.1)			
Refinement					
Resolution (Å)		40-2.80			
No. of reflections		320,814			
$R_{\text{work}}/R_{\text{free}}$		16.8/21.0			
No. amino acid residues		8,681			
No. of atoms					
Protein		68,820			
Ligand/ion		24			
Water		214			
B-factors (Å ²)					
Protein		81.1			
Ligand/ion		130.4			
Water		53.8			
R.m.s deviations					
Bond lengths (Å)		0.010			
Bond angles (°)		1.26			
Ramachandran [§]					
Preferred/allowed/disallowed (%)		94.5/4.6/0.9			

b	Crystal V	Peak	Inflection	Remote
Data collection*				
Wavelength (Å)		1.28288	1.28382	1.29639
Cell dimensions				
a, b, c (Å)		122.0, 139.2, 209.1		
α, β, γ (°)		108.3, 95.1, 94.2		
Resolution (Å)		40-3.5 (3.59-3.50) [†]	40-3.5 (3.59-3.50)	40-3.5 (3.59-3.50)
R_{sym}		15.1 (54.3)	17.0 (74.1)	21.3 (99.8)
$I/\sigma I$		16.54 (6.0)	8.14 (2.0)	6.53 (1.4)
Completeness (%)		98.0 (97.9)	97.2 (95.5)	97.1 (95.5)
Redundancy		9.8 (9.9)	5.0 (5.0)	5.0 (4.9)
$CC_{(1/2)}^{\ddagger}$ (%)		99.7 (94.5)	99.0 (67.9)	98.5 (47.3)

a, Diffraction data and refinement statistics. **b**, Multiwavelength anomalous diffraction data collection statistics.

* Diffraction data were collected on a Pilatus 6M detector (**a**) at beamline X10SA of the Swiss Light Source and ID23-1 and ID29 of the European Synchrotron Radiation Facility, and (**b**) at beamline X06SA of the Swiss Light Source. Diffraction data were processed with XDS². In **a**, data were collected at 0.25° or 0.1° increments (180° or 360° in total, respectively) at a wavelength of 1.0 Å. Data sets of four crystals were merged. The R_{pim} values were 0.072 (overall, 40–2.8 Å), 0.026 (40–15.3 Å) and 0.659 (2.85–2.80 Å). In **b**, data were collected in 0.1° increments (1080° in total). As a high-energy remote data set, the data given in **a** were used.

[†] Numbers in parentheses refer to the highest resolution shell.

[‡] $CC_{1/2}$ = percentage of correlation between intensities from random half-data sets⁶².

[§] Calculated with MolProbity⁶³.

A uniform metal distribution in the intergalactic medium of the Perseus cluster of galaxies

Norbert Werner^{1,2}, Ondrej Urban^{1,2,3}, Aurora Simionescu^{1,2,4} & Steven W. Allen^{1,2,3}

Most of the metals (elements heavier than helium) produced by stars in the member galaxies of clusters currently reside within the hot, X-ray-emitting intra-cluster gas. Observations of X-ray line emission from this intergalactic medium have suggested a relatively small cluster-to-cluster scatter outside the cluster centres^{1,2} and enrichment with iron out to large radii^{3–5}, leading to the idea that the metal enrichment occurred early in the history of the Universe³. Models with early enrichment predict a uniform metal distribution at large radii in clusters, whereas those with late-time enrichment^{6,7} are expected to introduce significant spatial variations of the metallicity. To discriminate clearly between these competing models, it is essential to test for potential inhomogeneities by measuring the abundances out to large radii along multiple directions in clusters, which has not hitherto been done. Here we

report a remarkably uniform iron abundance, as a function of radius and azimuth, that is statistically consistent with a constant value of $Z_{\text{Fe}} = 0.306 \pm 0.012$ in solar units out to the edge of the nearby Perseus cluster. This homogeneous distribution requires that most of the metal enrichment of the intergalactic medium occurred before the cluster formed, probably more than ten billion years ago, during the period of maximal star formation and black hole activity.

Between 2009 and 2011, we obtained a total of 84 observations of the Perseus cluster with the Suzaku X-ray satellite, as a 'key project' for that mission. The pointings covered eight azimuthal directions from the cluster centre out to an offset angle of 2° , with a total exposure time of over one million seconds. We have analysed the data from all three functioning X-ray imaging spectrometers, extracting spectra from annuli centred on the cluster centre. We modelled the spectra from

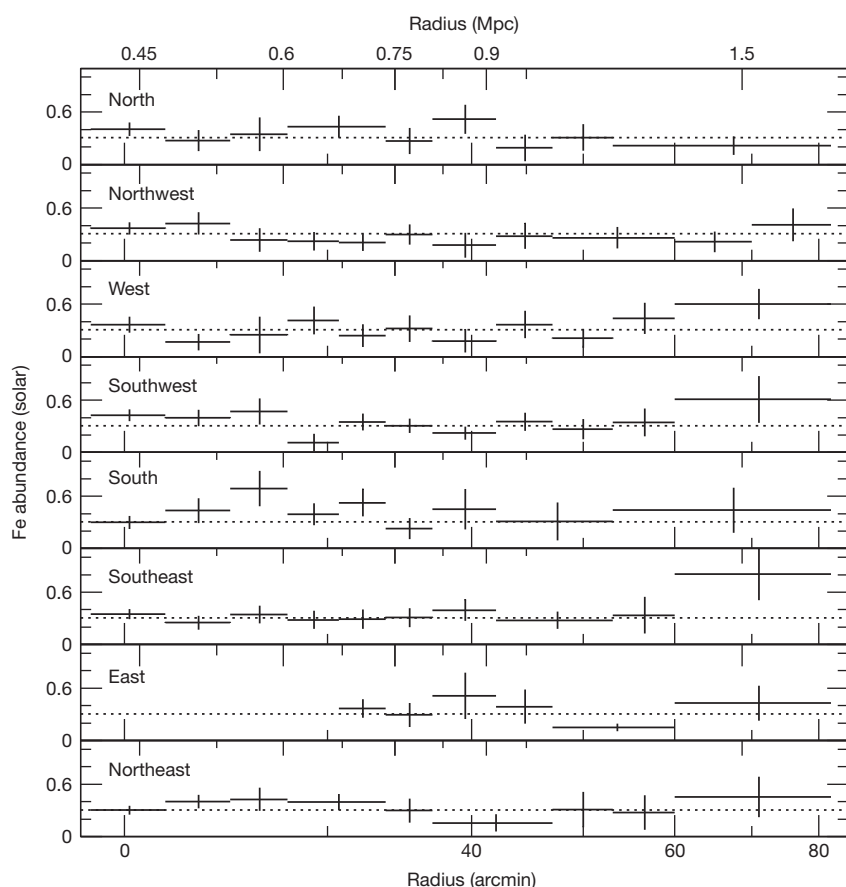


Figure 1 | Iron abundance profiles expressed in solar units¹⁰ (that is, as multiples of the solar abundance) measured along eight different directions in the Perseus Cluster. At the temperature of the Perseus Cluster, the iron abundance measurement is driven by the Fe–K lines. Compared to the Fe–L lines, these lines are significantly less prone to systematic biases that would be caused by the presence of multi-temperature gas, if present. The spectral extraction regions span a relatively small range in radius and azimuth, where the plasma can reasonably be assumed to have a single uniform temperature. There is no evidence for a bias in the temperature measurements caused by gas clumping⁸. We have verified that spectral fits that exclude the Fe–L line complex (0.8–1.5 keV energy range) give results consistent with those in the full 0.7–7-keV band. Collisional ionization and electron-ion equilibrium are likely to hold out to r_{200} in the Perseus cluster, where the equilibration timescales are only about 3.5×10^8 years and 7.3×10^8 years, respectively. Within a radius of around 400 kpc $\approx 0.2r_{200}$ (not shown here) metal enrichment is strongly influenced by the brightest cluster galaxy and the metallicity is centrally peaked^{4,9}. Along the eastern direction, large-scale sloshing motions are present and are uplifting the ICM from smaller radii out to $r \approx 650$ kpc (ref. 14). Therefore we also do not present results for $r < 650$ kpc in this direction. Beyond these radii, the iron abundance measurements are consistent with a radially and azimuthally constant value of $Z_{\text{Fe}} = 0.306 \pm 0.012$ solar, indicated by the dotted lines. Expressed in units of an older, and historically more commonly used, solar iron abundance value²⁹, our best-fit constant iron abundance is $Z_{\text{Fe}} = 0.212 \pm 0.008$ solar. The plotted error bars are 68% confidence intervals.

¹Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, California 94305-4085, USA. ²Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, California 94305-4060, USA. ³SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. ⁴Institute of Space and Astronautical Science (ISAS), JAXA, 3-1-1 Yoshinodai, Chuo-ku, Sagami-hara, Kanagawa 252-5210, Japan.

each of the 76 independent regions as a single-temperature thermal plasma in collisional ionization equilibrium, with the temperature, iron abundance and spectrum normalization included as free parameters⁸.

The measured radial and azimuthal variation of the iron abundance of the hot intracluster medium (ICM) out to the edge of the Perseus cluster along the eight different directions is presented in Fig. 1. We define this edge as r_{200} , the radius within which the mean enclosed mass density of the cluster is 200 times the critical density of the Universe at the cluster redshift (for the Perseus Cluster $r_{200} = 1.8$ Mpc, corresponding to 82 arcmin; ref. 4). We have tested the statistical uniformity of the measured iron abundance at radii $r > 400$ kpc (that is, beyond the central metallicity peak associated with the brightest cluster galaxy^{4,9}) by modelling the data from all radii and azimuths with a constant abundance. The measured chi-square value of 65.8 for 75 degrees of freedom is consistent with the null hypothesis of a constant iron abundance with a value of $Z_{\text{Fe}} = 0.306 \pm 0.012$ times the solar abundance (68% confidence limit; we adopt a solar iron abundance of 3.24×10^{-5} relative to hydrogen by number¹⁰). This abundance is also consistent with the mean value measured at intermediate radii (about $0.2r_{200}$ – $0.5r_{200}$) for a sample of 48 clusters previously observed with XMM-Newton². The iron abundance values measured at large radii in the less-massive Virgo cluster are lower, although these are probably biased by the multi-temperature structure of the gas in that system⁵. The azimuthally averaged iron abundance profile for the Perseus cluster is shown in Fig. 2.

Enrichment scenarios in which the ICM is polluted with metals after the formation of clusters predict a non-uniform metal distribution, which is inconsistent with the observations presented here. Ram-pressure stripping¹¹ would result in a negative gradient in the metal abundance profiles out to large radii, as well as significant azimuthal variations, with higher metal abundances expected along directions connecting to surrounding large-scale structure filaments¹². Furthermore, metals ejected by galaxies at later times should roughly follow the distribution of galaxies, producing an approximately constant ratio of metal mass to light, which is also inconsistent with observations¹³. Although large-scale sloshing motions¹⁴ may be able to mix metals to some degree, the strong gradients in the entropy

distributions of cluster atmospheres¹⁵, including that of Perseus^{4,8}, make the ICM convectively stable, prohibiting the efficient mixing of metals that are initially non-uniformly distributed across large radial ranges.

The observed uniform iron abundance distribution at large radii in the Perseus cluster requires early enrichment of the intergalactic gas in the proto-cluster environment. This enrichment was probably driven primarily by galactic winds¹⁶—energetic outflows of metal-enriched gas—which are expected to be strongest at epochs around the peak of star formation¹⁷ and active galactic nuclei activity¹⁸ (redshifts $z \approx 2$ – 3 , or lookback times of 10–12 billion years). The combined energy of supernova explosions¹⁶ and active galactic nuclei¹⁹ must have been strong enough to expel most of the metals from the galaxies at early times, and to enrich and mix the intergalactic gas. This gas was later accreted by clusters and virialized (increasing its entropy through shock heating) to form the present ICM.

The total iron mass in the ICM of the Perseus cluster, calculated from the measured iron abundance and ICM density profiles⁸, is about 50 billion solar masses, with about 60% of this mass residing beyond $0.5r_{200}$. The dominant fraction of this iron was probably supplied by type Ia supernovae, which are thought to have produced 60–90% of iron in the Perseus cluster¹³, depending on the assumed supernova yields. On the basis of our measurements, we estimate that at least 40 billion type Ia supernovae contributed to the chemical enrichment of the proto-cluster environment that later formed the Perseus cluster. For such a large iron mass to be expelled by the winds from the galaxies at early times, a significant fraction of these supernovae must have exploded shortly after the epoch of peak star formation. This is consistent with recent findings based on type Ia supernovae delay-time distributions^{20,21}, which imply that a large fraction of such supernovae explode less than about 5×10^8 years after the formation of the progenitor binary system (prompt type Ia supernovae). Type Ia supernovae with longer delay times will continue contributing to the enrichment of the ICM after clusters virialize, and are expected to be partly responsible for the metallicity peaks surrounding the brightest cluster galaxies in the centres of many clusters²².

A unique prediction of the early enrichment scenario is that essentially all galaxy clusters with masses comparable to the Perseus cluster should have homogeneous iron abundance distributions of about one-third of the solar abundance at large radii. Also, contrary to some initial findings^{6,7}, there should be no substantial redshift evolution in the ICM metallicity outside the central regions of clusters, out to $z \approx 2$. The presence and strength of such evolution are a matter of ongoing debate^{2,23,24}. The observed large iron abundance of the high-entropy gas in the outskirts of the Perseus cluster is also consistent with the idea that the highest-energy cosmic rays (above a few 10^{18} eV) may be primarily iron nuclei²⁵ accelerated by cluster formation shocks²⁶. Additionally, because much of the metal-rich ICM seen at large radii has been accreted from the surrounding large-scale structure filaments, our model predicts that the tenuous warm-hot intergalactic medium permeating the cosmic web, in which up to half of the baryons in the Universe currently reside^{27,28}, is likely to be substantially enriched in metals. The same is true for the hot circumgalactic gas accreted by galaxies from the intergalactic medium.

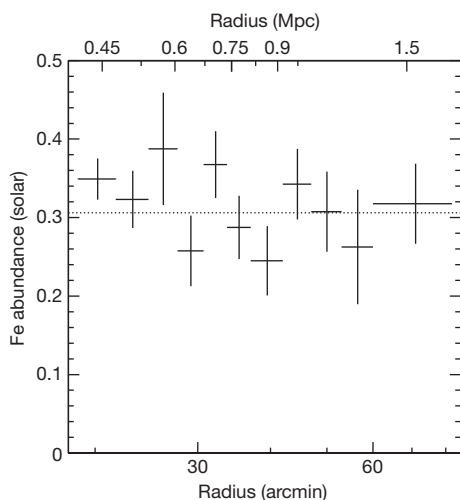


Figure 2 | Azimuthally averaged iron abundance profile for the Perseus cluster. The profile was determined by simultaneous fitting of the spectra at a given radius for all eight azimuthal directions. Although the iron abundance was assumed to be the same in all directions in the fits, the temperatures and spectral normalizations were allowed to vary independently along the different azimuths. The dotted line shows the best-fit constant iron abundance value of $Z_{\text{Fe}} = 0.306$ solar. The plotted error bars (vertical lines) are 68% confidence intervals. Our iron abundance values are lower than the value measured in the 650–1,100-kpc range with XMM-Newton¹³; however, the XMM-Newton measurement suffers from significant systematic uncertainties caused by the high and variable particle background of that instrument compared to that of Suzaku.

Received 1 June; accepted 9 September 2013.

1. Matsushita, K. Radial profiles of Fe abundance in the intracluster medium of nearby clusters observed with XMM-Newton. *Astron. Astrophys.* **527**, 134–146 (2011).
2. Leccardi, A. & Molendi, S. Radial metallicity profiles for a large sample of galaxy clusters observed with XMM-Newton. *Astron. Astrophys.* **487**, 461–466 (2008).
3. Fujita, Y. et al. High metallicity of the X-ray gas up to the virial radius of a binary cluster of galaxies: evidence of galactic superwinds at high-redshift. *Publ. Astron. Soc. Jpn* **60**, 343–349 (2008).
4. Simionescu, A. et al. Baryons at the edge of the X-ray-brightest galaxy cluster. *Science* **331**, 1576–1579 (2011).
5. Urban, O. et al. X-ray spectroscopy of the Virgo cluster out to the virial radius. *Mon. Not. R. Astron. Soc.* **414**, 2101–2111 (2011).

6. Balestra, I. *et al.* Tracing the evolution in the iron content of the intra-cluster medium. *Astron. Astrophys.* **462**, 429–442 (2007).
7. Maughan, B. J., Jones, C., Forman, W. & Van Speybroeck, L. Images, structural properties, and metal abundances of galaxy clusters observed with Chandra ACIS-I at $0.1 < z < 1.3$. *Astrophys. J.* **174**, 117–135 (2008).
8. Urban, O. *et al.* Azimuthally resolved X-ray spectroscopy of the Perseus cluster out to its edge. *Mon. Not. R. Astron. Soc.* (submitted); <http://arxiv.org/abs/1307.3592> (2013).
9. De Grandi, S., Ettori, S., Longhetti, M. & Molendi, S. On the iron content in rich nearby clusters of galaxies. *Astron. Astrophys.* **419**, 7–18 (2004).
10. Feldman, U. Elemental abundances in the upper solar atmosphere. *Phys. Scr.* **46**, 202–220 (1992).
11. Gunn, J. E. & Gott, J. R. III. On the infall of matter into clusters of galaxies and some effects on their evolution. *Astrophys. J.* **176**, 1–19 (1972).
12. Domainko, W. *et al.* Enrichment of the ICM of galaxy clusters due to ram-pressure stripping. *Astron. Astrophys.* **452**, 795–802 (2006).
13. Matsushita, K., Sakuma, E., Sasaki, T., Sato, K. & Simionescu, A. Metal-mass-to-light ratios of the Perseus cluster out to the virial radius. *Astrophys. J.* **764**, 147–156 (2013).
14. Simionescu, A. *et al.* Large-scale motions in the Perseus galaxy cluster. *Astrophys. J.* **757**, 182–187 (2012).
15. Voit, G. M., Kay, S. T. & Bryan, G. L. The baseline intracluster entropy profile from gravitational structure formation. *Mon. Not. R. Astron. Soc.* **364**, 909–916 (2005).
16. De Young, D. S. On the origin and evolution of iron-enriched gas in clusters of galaxies. *Astrophys. J.* **223**, 47–55 (1978).
17. Madau, P. *et al.* High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to $z \sim 4$. *Mon. Not. R. Astron. Soc.* **283**, 1388–1404 (1996).
18. Brandt, W. N. & Hasinger, G. Deep extragalactic X-ray surveys. *Annu. Rev. Astron. Astrophys.* **43**, 827–859 (2005).
19. Fabjan, D. *et al.* Simulating the effect of active galactic nuclei feedback on the metal enrichment of galaxy clusters. *Mon. Not. R. Astron. Soc.* **401**, 1670–1690 (2010).
20. Mannucci, F., Della Valle, M. & Panagia, N. Two populations of progenitors for Type Ia supernovae? *Mon. Not. R. Astron. Soc.* **370**, 773–783 (2006).
21. Maoz, D., Mannucci, F. & Brandt, T. D. The delay-time distribution of Type Ia supernovae from Sloan II. *Mon. Not. R. Astron. Soc.* **426**, 3282–3294 (2012).
22. Böhringer, H., Matsushita, K., Churazov, E., Finoguenov, A. & Ikebe, Y. Implications of the central metal abundance peak in cooling core clusters of galaxies. *Astron. Astrophys.* **416**, L21–L25 (2004).
23. Baldi, A. *et al.* An XMM-Newton spatially-resolved study of metal abundance evolution in distant galaxy clusters. *Astron. Astrophys.* **537**, 142–153 (2012).
24. Andreon, S. The enrichment history of the intracluster medium: a Bayesian approach. *Astron. Astrophys.* **546**, 6–17 (2012).
25. Abraham, J. *et al.* Measurement of the depth of maximum of extensive air showers above 10^{18} eV. *Phys. Rev. Lett.* **104**, 091101 (2010).
26. Norman, C. A., Melrose, D. B. & Achterberg, A. The origin of cosmic rays above $10^{18.5}$ eV. *Astrophys. J.* **454**, 60–68 (1995).
27. Cen, R. & Ostriker, J. P. Where are the baryons? *Astrophys. J.* **514**, 1–6 (1999).
28. Davé, R. *et al.* Baryons in the warm-hot intergalactic medium. *Astrophys. J.* **552**, 473–483 (2001).
29. Anders, E. & Grevesse, N. Abundances of the elements—meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).

Acknowledgements We are grateful for discussions with other members of the Perseus cluster Suzaku Key project collaboration, as well as with Y. Lu, P. Simeon and R. Blandford. This work was supported by the Suzaku grants NNX09AV64G and NNX10AR48G, the NASA ADAP grant NNX12AE05G, and by the US Department of Energy under contract number DE-AC02-76SF00515.

Author Contributions N.W. led the writing of the manuscript. O.U. reduced and analysed the data. A.S. and N.W. contributed to the data analysis. S.W.A. contributed to the writing of the manuscript and is the principal investigator of the Suzaku Key Project data. All authors discussed all results, developed the interpretation and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.W. (norbertw@stanford.edu).

Minimal-excitation states for electron quantum optics using levitons

J. Dubois^{1*}, T. Jullien^{1*}, F. Portier¹, P. Roche¹, A. Cavanna², Y. Jin², W. Wegscheider³, P. Roulleau¹ & D. C. Glattli¹

The on-demand generation of pure quantum excitations is important for the operation of quantum systems, but it is particularly difficult for a system of fermions. This is because any perturbation affects all states below the Fermi energy, resulting in a complex superposition of particle and hole excitations. However, it was predicted nearly 20 years ago^{1–3} that a Lorentzian time-dependent potential with quantized flux generates a minimal excitation with only one particle and no hole. Here we report that such quasiparticles (hereafter termed levitons) can be generated on demand in a conductor by applying voltage pulses to a contact. Partitioning the excitations with an electronic beam splitter generates a current noise that we use to measure their number. Minimal-excitation states are observed for Lorentzian pulses, whereas for other pulse shapes there are significant contributions from holes. Further identification of levitons is provided in the energy domain with shot-noise spectroscopy, and in the time domain with electronic Hong–Ou–Mandel noise correlations^{4–8}. The latter, obtained by colliding synchronized levitons on a beam splitter, exemplifies the potential use of levitons for quantum information: using linear electron quantum optics⁹ in ballistic conductors, it is possible to imagine flying-qubit^{10,11} operation in which the Fermi statistics are exploited^{12–14} to entangle synchronized electrons emitted by distinct sources^{15–18}. Compared with electron sources based on quantum dots^{19–21}, the generation of levitons does not require delicate nanolithography, considerably simplifying the circuitry for scalability. Levitons are not limited to carrying a single charge, and so in a broader context n -particle levitons could find application in the study of full electron counting statistics^{22–25}. But they can also carry a fraction of charge if they are implemented in Luttinger liquids³ or in fractional quantum Hall edge channels²⁶; this allows the study of Abelian and non-Abelian quasiparticles in the time domain. Finally, the generation technique could be applied to cold atomic gases^{27,28}, leading to the possibility of atomic levitons.

A non-trivial means of generating a time-resolved elementary excitation from the ground state of a system of fermions was given in a series of theoretical papers by Levitov and co-workers^{1–3}. The difficulty with fermions stems from the fact that any applied potential acts on all the electrons filling the Fermi sea—the cooperative effect is unlikely to converge towards a single elementary excitation. A well-known example is the Anderson orthogonality catastrophe²⁹, which makes the ground state of fermions with an impurity potential orthogonal to that without the impurity. A dynamical analogue was found by Lee and Levitov³⁰, who considered a voltage pulse, $V(t)$, applied to a quantum conductor. After the pulse, a charge of $Q = e \Delta\varphi / 2\pi$ per electronic quantum mode is transferred through the conductor, where $\Delta\varphi = \varphi(+\infty) - \varphi(-\infty)$, $\varphi(t) = (2\pi/h) \int_{-\infty}^t eV(t') dt'$ is the time-dependent phase acquired by electrons, e is the electron charge and h is Planck's constant. For non-integer charge, the final many-body state strongly differs from the initial one and contains a diverging number of electrons and holes created above and below the Fermi energy, E_F , which is a signature of the orthogonality

catastrophe. Remarkably, the catastrophe vanishes for integer charge $Q = en$ (that is, when $\Delta\varphi$ is a multiple of 2π). However, the charges transferred may still be accompanied by a finite amount of neutral excitation. Indeed, after experiencing the voltage pulse all electrons in the Fermi sea are in a superposition of states with energies shifted both upwards and downwards. A positive energy shift of the states initially filled below E_F results in electron-like excitations, whereas a negative shift of the empty states above E_F results in holes production.

Finding the condition to transfer n charges with no extra excitations is the second remarkable result of refs 1–3. For a Lorentzian voltage pulse of width $2W$ generating a unit charge

$$eV(t) = \frac{h}{\pi W} \frac{1}{1 + (t/W)^2} \quad (1)$$

a single electron is transferred free of hole excitations. Indeed, for $V(t)$ given by equation (1), $e^{-i\varphi(t)} = (t + iW)/(t - iW)$ has no pole in the lower complex plane. The probability of an electron having its energy shifted by $\Delta\varepsilon$, $P(\Delta\varepsilon) = \left| \int_{-\infty}^{+\infty} e^{-i\varphi(t)} e^{i\Delta\varepsilon t/\hbar} dt / \sqrt{2\pi\hbar} \right|^2$, where $\hbar = h/2\pi$, vanishes for negative $\Delta\varepsilon$ and only a positive energy boost is given to the Fermi sea. The wavefunction associated to the resulting time-resolved minimal excitation is schematically depicted in Fig. 1a. It shows an exponential decrease in energy, starting right above the Fermi surface, and it is Lorentzian in shape in the space or time domain. This new single-particle state has all the properties of an elementary excitation³. We call it a leviton because, like solitons, n levitons of the same charge sign but arbitrary width and creation time can be combined at will, even with an overlap.

Here we report the physical implementation of levitons in a ballistic one-dimensional conductor and their identification using shot noise. A narrow constriction, called a quantum point contact (QPC), is defined in the middle of a two-dimensional electron gas (2DEG) by applying a negative voltage to lateral gates (Fig. 1b). A single electron channel forms whose transmission, D , is controlled by the gate voltage. Two ohmic contacts, one on either side of the constriction, connect the 2DEG to a pulse generator through broadband, 40-GHz transmission lines (Fig. 1c). The shortest Lorentzian pulses are 30-ps wide for a repetition period $T = 1/\nu = 166$ ps. These timescales are smaller than the thermal time, $\hbar/k_B T_e \approx 220$ ps (k_B , Boltzmann's constant), for the lowest electronic temperature achieved, $T_e = 35$ mK.

Following the original proposal^{1–3}, the number of excitations generated by a pulse can be obtained from the low-frequency current noise resulting from the partitioning of all excitations by a beam splitter, here the QPC set at transmission $D < 1$. For periodic pulses, each generating n electrons, the smallest expected noise is $S_I = 2(2e^2/h)n\hbar\nu D(1-D) = S_I^0 n$. This is the shot noise that would occur if the periodic pulses were replaced by the d.c. voltage $\langle V(t) \rangle = V_{d.c.} = n\hbar\nu/e$. However, owing to the a.c. voltage component, $V_{a.c.}(t) = V(t) - V_{d.c.}$, in general N_e electrons and $N_h = N_e - n$ holes per pulse may be generated. Their number is linked to the probability, P_l , of electrons having their energy shifted by l quanta,

¹Nanoelectronics Group, Service de Physique de l'Etat Condensé, IRAMIS/DSM (CNRS URA 2464), CEA Saclay, F-91191 Gif-sur-Yvette, France. ²CNRS/Univ Paris Diderot (Sorbonne Paris Cité), Laboratoire de Photonique et de Nanostructures (LPN), route de Nozay, 91460 Marcoussis, France. ³Laboratory for Solid State Physics, ETH Zürich, CH-8093 Zürich, Switzerland.

*These authors contributed equally to this work.

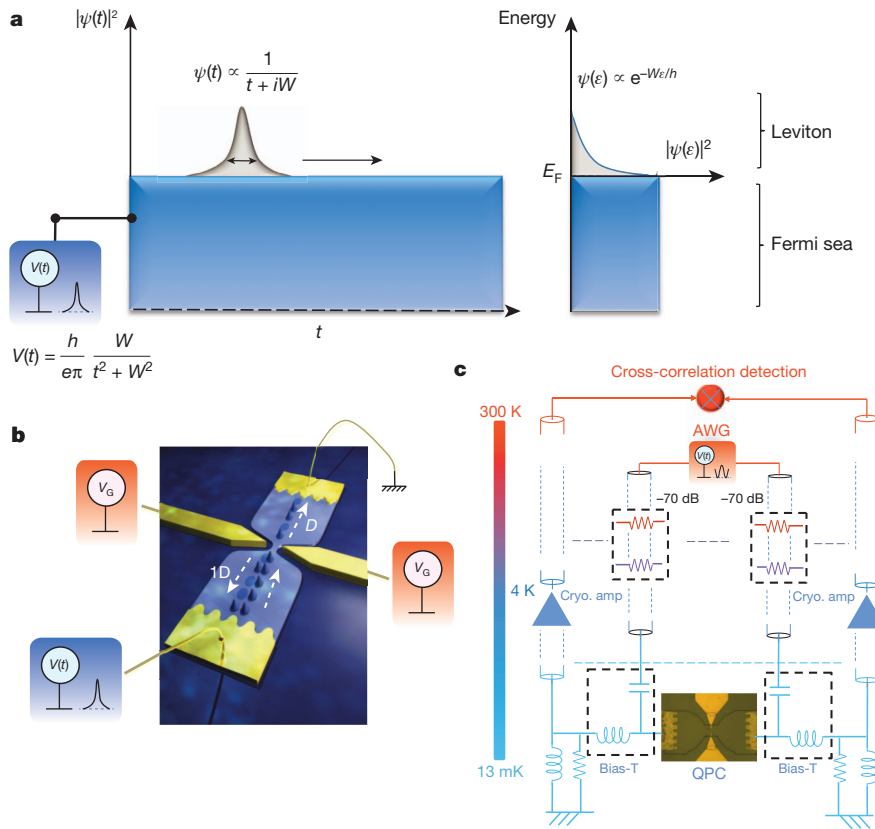


Figure 1 | Levitons and the principle of their experimental detection. **a**, Single-particle leviton: schematic picture of the wavefunction, ψ , in the time and energy domains. **b**, Principle of experiment: voltage pulses, $V(t)$, on the contact of a 2DEG generate charge pulses that are partitioned by a QPC. The lateral-gate voltage, V_G , controls the transmission, D , of the one-dimensional (1D) electronic mode formed at the QPC. **c**, Voltage pulses are sent from an arbitrary wave generator (AWG) to the ohmic contacts of the sample via 40-GHz transmission lines. Bias-Ts separate the high-frequency components from the d.c. voltage bias and the detected low-frequency current noise. The latter is converted into voltage fluctuations. A computer performs fast-Fourier-transformation cross-correlations after cryogenic amplification and fast acquisition.

that is, by $lh\nu$, when travelling from the left contact to the QPC. From ref. 31, $N_e = \sum_{l \geq -n} (l+n)P_l$ and $N_h = \sum_{l < -n} |l+n|P_l$, where $P_l = \left| \int_0^T dt/T e^{-i\varphi(t)} e^{i(l+n)vt} \right|^2$. The total noise is $S_I = S_I^0(N_h + N_e) \geq S_I^0 n$. Its minimization at constant current $I = ev(N_e - N_h) = evn$ gives $N_h = 0$ for positive n (and $N_e = 0$ for negative n). The absence of holes implies that $P_{l < -n} = 0$, extending to the periodic case the condition $P(\Delta\epsilon) = 0$ for $\Delta\epsilon < 0$ for a single pulse. This is satisfied only by Lorentzian pulses creating n -particle levitons. By contrast, sinusoidal pulses always carry particle-hole excitations because $P_l = J_l^2(l) \neq 0$ for all l , where J_l is a Bessel function. These two different cases create the qualitatively different energy distributions $\tilde{f}(\epsilon) = \sum_{l=-\infty}^{+\infty} P_l f(\epsilon - lh\nu)$, where $f(\epsilon) = 1 + \exp((\epsilon - E_F)/k_B T_e)$ is the equilibrium energy distribution.

Our measurements are presented in terms of the excess noise, ΔS_I , obtained by subtracting the noise with $V_{a.c.}$ 'off' from the noise with $V_{a.c.}$ 'on'. At zero temperature, ΔS_I measures the excess particle number, $\Delta N_{eh} = \Delta S_I/S_I^0 = N_e + N_h - n = 2N_h$, which can be written^{24,31}

$$\Delta N_{eh} = \sum_l |l+n|P_l - |n|$$

and vanishes for levitons. In the case of non-integer charge pulses (with q replacing n), as a consequence of the orthogonality catastrophe³⁰ both electron and hole excitations are required, making ΔN_{eh} finite even for Lorentzian pulses. Remarkably, all pulse shapes give a local minimum at integer charge^{24,31} (Supplementary Fig. 3). At finite temperature, the excess noise normalized to S_I^0 becomes³¹

$$\Delta N_{eh} = \sum_{l=-\infty}^{+\infty} (l+q) \coth\left(\frac{l+q}{2\theta_e}\right) P_l - q \coth(q/2\theta_e) \quad (2)$$

This slightly overestimates the actual particle number even when $\theta_e = k_B T_e/h\nu \ll 1$. Indeed, an extra contribution (typically $\sim 2\theta_e P_{-n}$ for $q = n$; see Supplementary Information) comes from the partition noise of the thermal excitations emitted by the right-hand reservoir.

Their partition noise is promoted by the reduced occupation of the left-hand states when $V_{a.c.}$ is 'on' ($\tilde{f}(\epsilon \approx E_F) < 1$) and is inhibited by antibunching for $V_{a.c.}$ 'off' because then $\tilde{f}(\epsilon \approx E_F) = 1$.

We first report excess-noise measurements for different pulse shapes. Two similar samples and two different noise detection set-ups were used. Results for sample A using a 100–300-kHz frequency range and room-temperature amplifiers are presented in Supplementary Information. Results for sample B using cryogenic amplifiers and noise detection in a 450-kHz bandwidth around 2.5 MHz are presented here. The amplitude of $V_{a.c.}(t)$ for sinusoidal and square pulses is $V_{a.c.} = V_{d.c.} = qh\nu/e$. With $V_{a.c.}$ defined in the same way, the voltage for periodic Lorentzian pulses is $V(t) = V_{a.c.} \sinh(2\pi\eta)/2(\sinh(2\pi\eta)^2 + \sin(2\pi t/T)^2)$, where $\eta = W/T$.

Figure 2a shows ΔN_{eh} versus q for three pulse shapes. A clear hierarchy is visible. The square wave generates the largest noise, the Lorentzian generates the lowest and the sine wave is in between. For $q = 1$, the $3.7 \pm 1.2\%$ finite noise observed for the 4.8-GHz Lorentzian pulses is probably due to the thermal excitations of the right-hand reservoir, which are estimated to be $2\theta_e e^{-4\pi\eta} = 3.6\%$ ($W/T = 0.18$ and $\theta_e = 0.17$). By contrast, the expected $2\theta_e J_1^2(1) = 3.7\%$ thermal contribution for the 7.5-GHz sine wave cannot account for the 8.3% noise observed for $q = 1$: the sine wave shows clear hole excitation content that is not detectable for the Lorentzian. Figure 2b shows data for a sharper $W/T = 0.09$ Lorentzian pulse and higher-frequency sinusoidal pulses (the square-wave trace is repeated for scale comparison). The higher pulse energy makes thermal effects weaker, revealing the dynamical orthogonality catastrophe (DOC) suppression for integer charge by a noise minimum. For the Lorentzian pulse, the zero-temperature variations of ΔN_{eh} being asymmetrical around $q = 1$, thermal rounding shifts the minimum to $q \approx 1.4$ as expected (Supplementary Fig. 4). For sine waves, the almost symmetrical variations keep it centred near $q \approx 1$. However, at finite temperature a noise minimum is only an indirect signature of DOC suppression and in general cannot be used to find the exact charge for which an excitation minimum occurs. Indeed, we observe that for $q = 1$

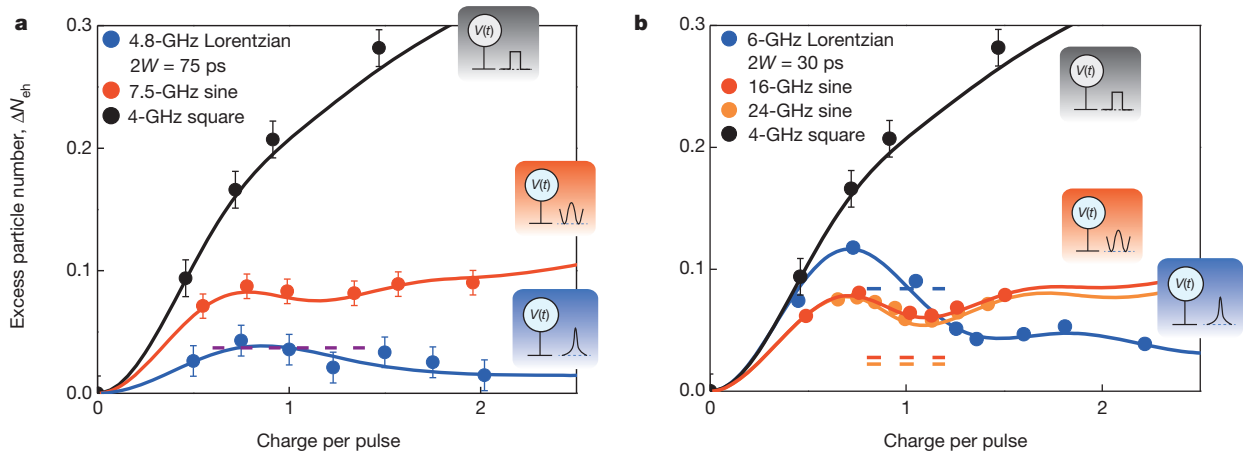


Figure 2 | Electron-hole excitation content of charge pulses and the dynamical orthogonality catastrophe. **a**, Excess particle number versus charge per pulse: dots are measurements; solid lines are comparisons with equation (2) including small heating effects. Square waves (black) give the largest noise, and Lorentzian pulses (blue) give the smallest ($W/T = 0.18$). The sine wave (red) is in between. The amount of extra noise around $q = 1$ generated by thermal excitations is $\sim 2(k_B T_e / \hbar \nu) e^{-4\pi W/T}$ for the Lorentzian and $\sim 2(k_B T_e / \hbar \nu) J_1(1)^2$ for the sinusoidal pulse. It is indicated as a short dashed purple line (both contributions are coincidentally equal). The thermal excitations account for the noise observed for Lorentzian pulses, as expected for a minimal excitation state, whereas for sinusoidal pulses the electron-hole excitation content gives larger noise. **b**, Same as **a** but using higher-frequency (16- and 24-GHz) sinusoidal pulses and a sharper, higher-frequency (6-GHz)

Lorentzian to observe the manifestation of the DOC³⁰. Because non-integer charges require many hole excitations, the larger noise is observed around $q = 0.6$ for both pulse types. The suppression of the DOC at integer charge is signalled by a noise minimum at $q \approx 1$ for the sinusoidal pulses. For the Lorentzian pulse, thermal averaging shifts this minimum to higher value, $q \approx 1.4$. However, it is actually for $q = 1$ that the number of excitations vanishes: indeed, the finite noise observed is entirely due to thermal excitations (blue dashed line). The red and orange dashed lines respectively indicate the thermal contributions to noise for the 16- and 24-GHz sinusoidal pulses (55 and 67 mK effective temperature) at $q = 1$ and reveal their finite hole content. All error bars are the s.e.m. calculated from the cross-correlation noise spectrum standard deviation in an 800-KHz bandwidth and after $n \approx 100,000$ acquisitions.

(not a noise minimum), our Lorentzian pulses contribute less than 1% to the excitations. The $9 \pm 1\%$ noise observed is again entirely due to thermal excitation partition noise: $\sim 2\theta_e e^{-4\pi\eta} = 8.6\%$ ($T_e = 39$ mK). For applications in quantum physics, it might be necessary to lower the amount of thermal excitations. A reduction by a factor of three seems possible by better filtering and improving the thermal conductance of the leads.

Finally, knowing the transmission, D , from conductance measurements, knowing $V_{a.c.}$ from calibration and knowing T_e from independent d.c. shot-noise measurements, and including small electron heating effects (Methods), theory³¹ (Fig. 2a, b, solid curves) and data show excellent agreement. In the following, we complete the identification of levitons from their signature in the energy and time domains.

Spectroscopic information³¹ can be extracted from shot noise. To understand this, consider the schematic diagrams in Fig. 3a, b, which are drawn for unit-charge pulses obtained for equal d.c. and a.c. voltages ($q = \alpha = 1$ in reduced units: $q = eV_{d.c.}/\hbar\nu$ and $\alpha = eV_{a.c.}/\hbar\nu$). Now we let the d.c. voltage vary, keeping the a.c. component constant. For a positive shift of $V_{d.c.}$, some hole excitations created in the left-hand side by the a.c. potential are promoted above the Fermi energy, E_F , of the right-hand reservoir and their contribution to noise vanishes. Similarly, for a negative shift of $V_{d.c.}$, some electron excitations on the left-hand side are below E_F . Their anti-bunching with electrons emitted from the right makes their noise contribution vanish. This gives direct information on the non-equilibrium energy distribution, $f(\epsilon)$. At zero temperature, the excess noise in particle number units is, for $q > 0$, $\Delta N(q) = \int_{-\infty}^{-q} (1 - \tilde{f}(\epsilon)) d\epsilon$, probing the hole energy distribution; for $q < 0$, it is $\Delta N(q) = \int_{-\infty}^{-q} \tilde{f}(\epsilon) d\epsilon$, probing electrons. For pulses carrying n charges, $\tilde{f}(\epsilon)$ is symmetric around $E_F + n\hbar\nu$ for symmetric sinusoidal and square pulses (Fig. 3a), but is asymmetric for asymmetric pulses such as rectangular pulses, the extreme case being for n -particle levitons where $\tilde{f}(\epsilon < E_F) = 0$ (Fig. 3b). This qualitative difference is reflected in the excess noise as observed for the sinusoidal and Lorentzian pulses (Fig. 3d, c), which respectively show symmetric and asymmetric noise variations with $V_{d.c.}$ for different a.c. amplitudes α . We note that

when q crosses the value α the noise observed in Fig. 3c, d corresponds to the excess particle number, $\Delta N_{eh}(q) = \Delta N(q = \alpha)$, displayed in Fig. 2b. Although thermal excitations make the noise larger for Lorentzian pulses than for sinusoidal pulses for $q = \alpha \approx 1$, the noise rapidly vanishes beyond the thermal energy ($q > \alpha + k_B T_e / \hbar\nu$), signalling the absence of holes, which is a hallmark of levitons. In contrast, the sine waves show a slower linear decrease. For instance, for $q = 1.6$ the noise for the Lorentzian with $\alpha = 1.05$ is less than 1.5%, whereas for the two sinusoidal pulses with $\alpha = 0.94$ and, respectively, 1.13 (average $\alpha = 1.03$) the noise, averaged for comparison, is still $\sim 2.3\%$.

Finally, to get time-domain information on the leviton wave packets, we closely follow the historical approach of Hong, Ou and Mandel⁴, who were interested in measuring the time width of single photon wave packets. The lack of quadratic detectors to perform time auto-correlation at so low an input level led them to consider the second-order coherence, $g_2(\tau) = |\langle \psi(x) | \psi(x - v_F \tau) \rangle|^2$, by colliding the idler and signal photons generated by parametric down-conversion of a laser source on a beam splitter. Indeed, the interference of the two indistinguishable particles makes the particle detection statistics dependent on their wavefunction overlap. After N_0 experiments, the particle number fluctuation is $\Delta N^2 = 2(1/4)N_0(1 \pm |\langle \psi(x) | \psi(x - v_F \tau) \rangle|^2)$, where the plus sign holds for bosons, the minus sign holds for fermions, τ is the time delay between particles and v_F is their velocity. For non-overlapping states at large τ , the fluctuations of two particles independently partitioned is found. For zero delay (full overlap), the bosonic bunching doubles the noise whereas the fermionic exclusion makes it vanish. Hong–Ou–Mandel experiments are now standard in quantum optics. With the use of electronic beam splitters, d.c. voltage sources have shown anti-bunching^{5–7} and $g_2(\tau)$ has been measured for energy-resolved electrons⁸.

To measure $g_2(\tau)$ for single-charge levitons, we apply Lorentzian pulses $V(t)$ and, respectively, $V(t + \tau)$ to the left- and right-hand contacts (Fig. 4a and b). This generates indistinguishable electrons whose two-particle interference reduces the partition shot noise from $2S_I^0$ to

$$S_I^{\text{HOM}} = 2S_I^0(1 - g_2(\tau))$$

This holds only for levitons, and not for other electron pulses³¹.

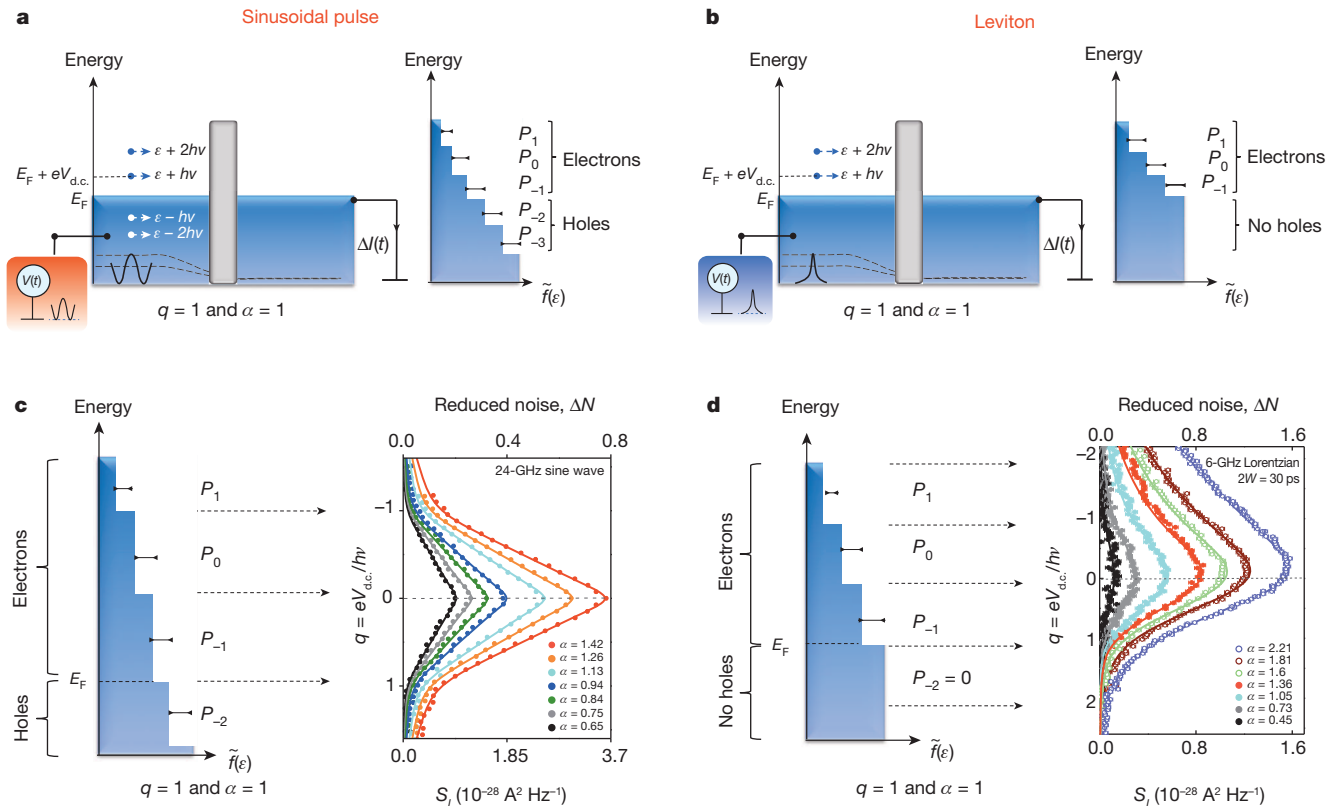


Figure 3 | energy distribution of electron-hole excitations: shot-noise spectroscopy. **a**, Electrons with energy ϵ enter the QPC barrier in a superposition of states with energy $\epsilon + \hbar\nu$ with probability P_i . Electrons and holes are partitioned, generating a current noise proportional to their number. $\tilde{f}(\epsilon)$ is the energy distribution for unit-charge sinusoidal pulses. It is symmetric with respect to the energy $E_F + eV_{d.c.}$. **b**, Same as **a** but for unit-charge Lorentzian pulses. $\tilde{f}(\epsilon)$ is asymmetrical because no holes are created, which is a hallmark of levitons. **c**, Shot-noise spectroscopy for 24-GHz sine waves with different amplitudes $V_{a.c.}$ (units, $\alpha = eV_{a.c.}/\hbar\nu$). The reduced excess noise, ΔN (dots), versus d.c. voltage in units of $q = eV_{d.c.}/\hbar\nu$ is plotted and compared with theory (solid lines). The QPC transmission is $D = 0.25$. The left-hand

diagram represents the distribution function. The noise derivative is proportional to $\tilde{f}(\epsilon)$ for $q < 0$ (that is, $\epsilon > E_F + eV_{d.c.}$) and to $-(1 - \tilde{f}(\epsilon))$ for $q > 0$ (see text). The symmetric noise curve reflects the symmetric distribution $\tilde{f}(\epsilon)$. For $q > \alpha$, the noise slowly linearly decreases as the hole excitations spread deep in energy below E_F . **d**, Same as **c** but for sharp, $W/T = 0.09$, 6-GHz Lorentzian pulses with various amplitudes α ($D = 0.19$). As expected, ΔN is strongly asymmetric. For $q > \alpha + k_B T_e/\hbar\nu$, in an energy domain beyond thermal excitations, the noise rapidly exponentially vanishes, signalling the absence of hole excitations. When q equals α , the reduced noise $\Delta N(q = \alpha)$ displayed in **c** and **d** equals the excess particle $\Delta N_{eh}(q)$ displayed in Fig. 2d for Lorentzian and sinusoidal pulses. The error bars are s.e.m. ($n \approx 100,000$ acquisitions).

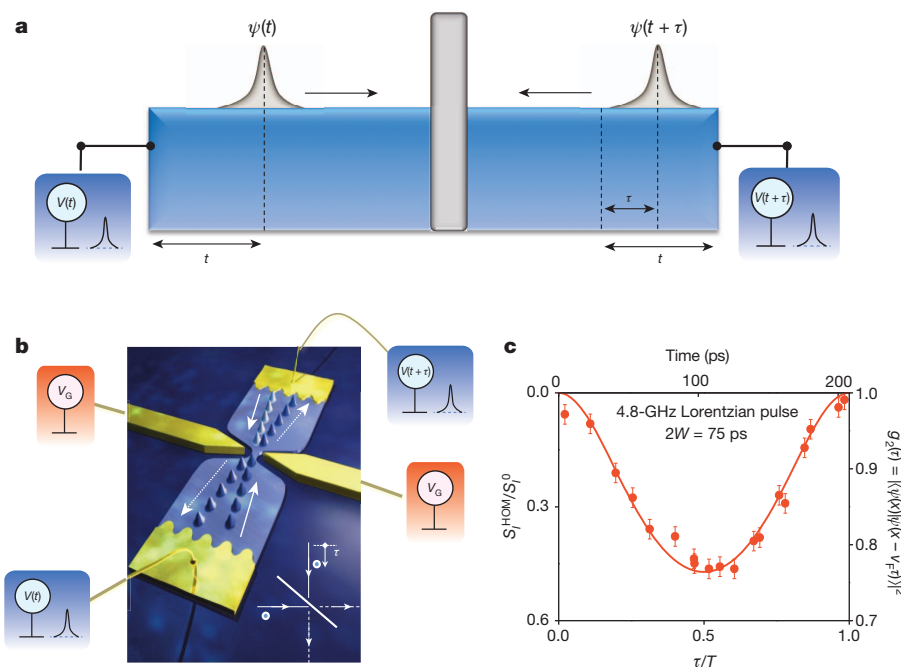


Figure 4 | Leviton wavefunction in the time domain. **a**, **b**, Principle of the Hong-Ou-Mandel experiment: time-shifted Lorentzian voltage pulses $V(t)$ and $V(t + \tau)$ applied to the left- and right-hand contacts generate trains of indistinguishable electrons that enter the two inputs of the QPC beam splitter. Their interference controls the statistics of the charge leaving the QPC. For single-charge levitons, $S_I^{\text{HOM}} = 2S_I^0(1 - |\langle \psi(x) | \psi(x - v_F \tau) \rangle|^2)$ (ref. 31). **c**, Hong-Ou-Mandel noise versus time delay for a 4.8-GHz Lorentzian with $W/T = 0.18$. The shot-noise scale on the left has been reversed to match the $g_2(\tau)$ scale on the right. The excellent agreement with theory (solid line) shows that the leviton wavefunction has the expected time dependence. The $\sim 100\%$ Hong-Ou-Mandel dip reveals the high symmetry of the right- and left-hand voltage pulses, which therefore provide indistinguishable electrons. The error bars are s.e.m. ($n \approx 100,000$ acquisitions).

Figure 4c shows $g_2(\tau)$ for 4.8-GHz Lorentzian pulses with $W/T = 0.18$. At $\tau = 0$, a near-unit g_2 proves the good symmetry between right- and left-hand radio-frequency voltage pulses, which therefore create indistinguishable excitations from the separate left- and right-hand contacts. Half way through the pulse, $g_2(\tau)$ is minimal. The excellent agreement of the measured $g_2(\tau)$ with theory (Fig. 4c, solid line) with no adjustable parameter demonstrates that leviton wave packets behave as expected in the time domain. A different variation is found for sinusoidal pulses, proving that the experiment can discriminate different pulse types (Supplementary Fig. 7). With our technique, Hong–Ou–Mandel noise correlations can be extended to n -particle levitons.

The on-demand generation of remarkable minimal-excitation fermionic states, which we call levitons, is demonstrated by applying Lorentzian voltage pulses to a quantum conductor. Shot-noise spectroscopy and electronic Hong–Ou–Mandel interference have probed their wavefunctions in the energy and time domains. Our technique already achieves an extra excitation of less than $\sim 1\%$ per leviton, opening a new field of quantum experiments.

METHODS SUMMARY

The samples are 2DEGs with electron mobilities of $2 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and densities of $\sim 1.4 \times 10^{15} \text{ m}^{-2}$ buried 100 nm under the surface of a GaAs/Ga(Al)As heterojunction. Wet-etching and lithography are used to pattern two $\sim 50\text{-}\mu\text{m}$ -long electron reservoirs separated by a constriction (Fig. 1). Low-resistance AuNiGe ohmic contacts connect the 2DEG to the external circuit. In the middle of the constriction, 300-nm-gap metallic lateral gates defined by electron-beam lithography are evaporated. Applying a negative gate voltage at low temperature depletes the 2DEG to form the QPC. For all measurements, a single orbital mode is transmitted. Conductance measurements provide its transmission via $G = D2e^2/h$. The samples are mounted on the 13 mK stage of a $\sim 400\text{-}\mu\text{W}$ Cryoconcept pulse-tube dilution refrigerator. A printed circuit board with two coplanar transmission lines designed using the CST Microwave Studio 3D electromagnetic field solver links the sample to two coaxial lines running from the top of the cryostat. Attenuators distributed at different temperatures reduce the thermal radiation. The overall set-up bandwidth is 40 GHz. A 24 GS s^{-1} Tektronix AWG7122B generator provides arbitrary pulse shapes. For sharper 6-GHz Lorentzian pulses, a combination of four harmonics is used. The current noise is converted into voltage fluctuations across two $\sim 2.5\text{-k}\Omega$ resistors in series with the sample. For the data presented in the main text, 450-kHz-bandwidth, 2.5-MHz tank circuits combined with home-made 0.2 nV Hz^{-1} cryogenic amplifiers are used. After further amplification and digitization, the cross-correlation voltage spectrum is calculated. For averaging over one minute, the radio-frequency on/radio-frequency off noise difference accuracy is $\sim 3 \times 10^{-30} \text{ A}^2 \text{ Hz}^{-1}$. The data are compared with theoretical expressions given in ref. 31. Self-heating effects are included by calculating the electronic temperature, T_e , from the Wiedemann–Franz law, considering the power, $G(V_{\text{d.c.}}^2 + V_{\text{a.c.}}^2/2)$, generated at the QPC and the finite lead and contact thermal resistance.

Received 26 July; accepted 18 September 2013.

Published online 23 October 2013.

- Levitov, L. S., Lee, H. & Lesovik, G. Electron counting statistics and coherent states of electric current. *J. Math. Phys.* **37**, 4845–4856 (1996).
- Ivanov, D. A., Lee, H. W. & Levitov, L. S. Coherent states of alternating current. *Phys. Rev. B* **56**, 6839–6850 (1997).
- Keeling, J., Klich, I. & Levitov, L. Minimal excitation states of electrons in one-dimensional wires. *Phys. Rev. Lett.* **97**, 116403 (2006).
- Hong, C. K., Ou, Z. Y. & Mandel, L. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
- Henny, M. *et al.* The fermionic Hanbury Brown and Twiss experiment. *Science* **284**, 296–298 (1999).
- Oliver, W. D., Kim, J., Liu, R. C. & Yamamoto, Y. Hanbury Brown and Twiss-type experiment with electrons. *Science* **284**, 299–301 (1999).

- Liu, R. C., Odom, B., Yamamoto, Y. & Tarucha, S. Quantum interference in electron collision. *Nature* **391**, 263–265 (1998).
- Bocquillon, E. *et al.* Coherence and indistinguishability of single electrons emitted by independent sources. *Science* **339**, 1054–1057 (2013).
- Ji, Y. *et al.* An electronic Mach–Zehnder interferometer. *Nature* **422**, 415–418 (2003).
- Bertoni, A., Bordone, P., Brunetti, R., Jacoboni, C. & Reggiani, S. Quantum logic gates based on coherent electron transport in quantum wires. *Phys. Rev. Lett.* **84**, 5912–5915 (2000).
- Yamamoto, M. *et al.* Electrical control of a solid-state flying qubit. *Nature Nanotechnol.* **7**, 247–251 (2012).
- Burkard, G., Loss, D. & Sukhorukov, E. V. Noise of entangled electrons: bunching and antibunching. *Phys. Rev. B* **61**, R16303–R16306 (2000).
- Beenakker, C. W. J., Emary, C., Kindermann, M. & van Velsen, J. L. Proposal for production and detection of entangled electron-hole pairs in a degenerate electron gas. *Phys. Rev. Lett.* **91**, 147901 (2003).
- Samuelsson, P., Sukhorukov, E. V. & Büttiker, M. Two-particle Aharonov–Bohm effect and entanglement in the electronic Hanbury Brown–Twiss setup. *Phys. Rev. Lett.* **92**, 026805 (2004).
- Ol'khovskaya, S., Splettstoesser, J., Moskalets, M. & Büttiker, M. Shot noise of a mesoscopic two-particle collider. *Phys. Rev. Lett.* **101**, 166802 (2008).
- Splettstoesser, J., Moskalets, M. & Büttiker, M. Two-particle nonlocal Aharonov–Bohm effect from two single-particle emitters. *Phys. Rev. Lett.* **103**, 076804 (2009).
- Haack, G., Moskalets, M., Splettstoesser, J. & Büttiker, M. Coherence of single-electron sources from Mach–Zehnder interferometry. *Phys. Rev. B* **84**, 081303 (2011).
- Sherkunov, Y. B., d'Ambrumenil, N., Samuelsson, P. & Büttiker, M. Optimal pumping of orbital entanglement with single-particle emitters. *Phys. Rev. B* **85**, 081108 (2012).
- Fève, G. *et al.* An on-demand coherent single-electron source. *Science* **316**, 1169–1172 (2007).
- Hermelin, S. *et al.* Electrons surfing on a sound wave as a platform for quantum optics with flying electrons. *Nature* **477**, 435–438 (2011).
- McNeil, R. P. G. *et al.* On-demand single-electron transfer between distant quantum dots. *Nature* **477**, 439–442 (2011).
- Levitov, L. S. & Lesovik, G. B. Charge distribution in quantum shot noise. *Pis'ma Z. Eksp. Fiz.* **58**, 225–230 (1993); *JETP Lett.* **58**, 230–235 (1993).
- Hassler, F., Lesovik, G. B. & Blatter, G. Effects of exchange symmetry on full counting statistics. *Phys. Rev. Lett.* **99**, 076804 (2007).
- Vanević, M., Nazarov, Y. V. & Belzig, W. Elementary events of electron transfer in a voltage-driven quantum point contact. *Phys. Rev. Lett.* **99**, 076601 (2007).
- Sherkunov, Y. B., Pratap, A., Muzykantskii, B. & d'Ambrumenil, N. Full counting statistics as the geometry of two planes. *Phys. Rev. Lett.* **100**, 196601 (2008).
- Jonckheere, T., Creux, M. & Martin, T. Time-controlled charge injection in a quantum Hall fluid. *Phys. Rev. B* **72**, 205321 (2005).
- Brantut, J. P. *et al.* Conduction of ultracold fermions through a mesoscopic channel. *Science* **337**, 1069–1071 (2012).
- Thywissen, J. H., Westervelt, R. M. & Prentiss, M. Quantum point contacts for neutral atoms. *Phys. Rev. Lett.* **83**, 3762–3765 (1999).
- Anderson, P. W. Infrared catastrophe in Fermi gases with local scattering potential. *Phys. Rev. Lett.* **18**, 1049–1051 (1967).
- Lee, H. W. & Levitov, L. Orthogonality catastrophe in a mesoscopic conductor due to a time-dependent flux. Preprint at <http://arxiv.org/abs/cond-mat/9312013> (1993).
- Dubois, J. *et al.* Integer and fractional charge Lorentzian voltage pulses analyzed in the framework of photon-assisted shot noise. *Phys. Rev. B* **88**, 085301 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements The ERC Advanced Grant 228273 MeQuaNo is acknowledged. We thank P. Jacques for technical help and P. Pari and P. Forget for support with cryogenics.

Author Contributions D.C.G. designed the project. J.D. fabricated sample A on a wafer provided by W.W., set up the radio-frequency and cryogenic systems with T.J., and, together with P. Rouleau, did the measurement and data analysis. The cryogenic amplifiers were made by P. Rouleau and T.J. F.P. helped in the early stages of the experiment and, together with P. Rouleau, T.J., P. Roche and D.C.G., wrote the paper. Sample B was provided by Y.J. on a wafer from A.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C.G. (christian.glatli@cea.fr).

Coupling a single electron to a Bose–Einstein condensate

Jonathan B. Balewski¹, Alexander T. Krupp¹, Anita Gaj¹, David Peter², Hans Peter Büchler², Robert Löw¹, Sebastian Hofferberth¹ & Tilman Pfau¹

The coupling of electrons to matter lies at the heart of our understanding of material properties such as electrical conductivity. Electron–phonon coupling can lead to the formation of a Cooper pair out of two repelling electrons, which forms the basis for Bardeen–Cooper–Schrieffer superconductivity¹. Here we study the interaction of a single localized electron with a Bose–Einstein condensate and show that the electron can excite phonons and eventually trigger a collective oscillation of the whole condensate. We find that the coupling is surprisingly strong compared to that of ionic impurities, owing to the more favourable mass ratio. The electron is held in place by a single charged ionic core, forming a Rydberg bound state. This Rydberg electron is described by a wavefunction extending to a size of up to eight micrometres, comparable to the dimensions of the condensate. In such a state, corresponding to a principal quantum number of $n = 202$, the Rydberg electron is interacting with several tens of thousands of condensed atoms contained within its orbit. We observe surprisingly long lifetimes and finite size effects caused by the electron exploring the outer regions of the condensate. We anticipate future experiments on electron orbital imaging, the investigation of phonon-mediated coupling of single electrons, and applications in quantum optics.

Charged impurities were very successfully used as probes of elementary excitations in the early studies of superfluidity in liquid helium. These applications span the interaction of ion impurities with phonons and rotons², the creation and study of vortex lattices by impurities³, and the coupling of electrons to surface ripplons⁴. Additionally they have been proposed for applications in quantum information⁵. The emergence of Bose–Einstein condensation of alkali atoms has renewed interest in impurity physics. Positively charged impurities in a Bose–Einstein condensate (BEC) were first created by Penning ionization of metastable atoms⁶ and photoionization⁷. Recently, single ion trapping⁸ has allowed the first study of the interaction between a single charged impurity and a BEC in a well controlled manner. In addition to bare losses by classical scattering, chemical reactions (both with the impurity⁹ as well as catalytic reactions¹⁰ mediated by the impurity) have been observed. Alternatively, the creation of neutral spin impurities in Fermi liquids offers tunable interaction strength with the bulk by means of a Feshbach resonance: the character of the interaction can be changed, from free impurities via a quasiparticle state (so-called Fermi polarons) to bound diatomic molecules^{11,12}.

For all such impurities, the interaction is inversely proportional to the reduced mass of the impurity and the bulk species¹³. Light impurities such as electrons are therefore, in general, better suited to obtaining strong coupling. However, electrons require an appropriate trapping potential, because even the tiniest electric field will lead to a considerable acceleration. The simplest trap for an electron offered by nature is a positively charged nucleus. We use highly excited Rydberg states, in which the electron is delocalized over regions up to several micrometres from the atomic core. At these distances the binding to the core is weak and the Rydberg electron becomes quasi-free, making it susceptible to interaction with its environment. Nonetheless, the binding to the ionic core is strong enough to provide sufficient trapping of the electron even in a strongly interacting environment. In fact, Rydberg absorption series

have been observed over a range of principal quantum numbers (n) around 30 at pressures reaching one atmosphere in 1934 (ref. 14). At such high densities, the spatial extent of the electron wavefunction is much larger than the mean interparticle distance, causing large interaction-induced line shifts and broadenings. The explanation of these effects led to the now well known Fermi pseudopotential¹⁵, which describes the short-range interaction of the quasi-free Rydberg electron at position r with neutral ground-state atoms at R :

$$V_{\text{pseudo}}(r, R) = \frac{2\pi\hbar^2 a}{m_e} \delta(r - R) \quad (1)$$

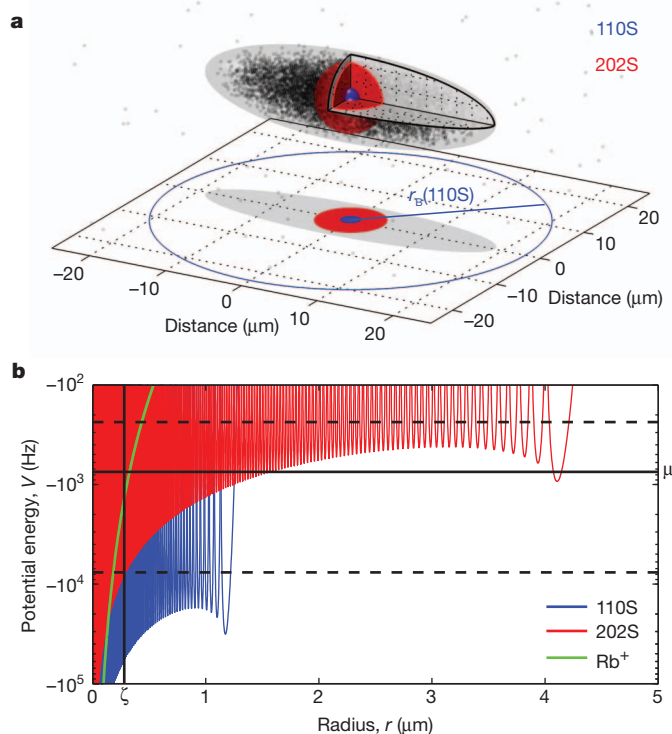


Figure 1 | Size comparison in the spatial and energy domains. **a**, Depending on the Rydberg state, the electron impurity is localized in different volumes in a BEC consisting of $N = 8 \times 10^4$ atoms. The sizes of the lowest (110S, blue) and highest Rydberg state (202S, red) under investigation are indicated. The densities of the BEC and the surrounding thermal cloud are to scale. The lower bound of the blockade radius r_B for the 110S state is denoted as a blue circle in the projection. The blockade radii for the higher Rydberg states are off the scale. **b**, The corresponding interaction potentials $V(r)$ (blue and red) from equation (2) are orders of magnitude stronger than the contribution of the positively charged Rydberg core Rb^+ (green) except for very small distances. The mean interaction strength (black dashed lines) can be set below and above the chemical potential $\mu = 745$ Hz of the condensate (horizontal black line) by choosing the Rydberg state. The healing length $\xi = 274$ nm of the condensate (vertical black line) is much smaller than the spatial extent of the electron wavefunction.

¹Physikalisches Institut, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany. ²Institut für Theoretische Physik III, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany.

Here the interaction strength is fully characterized by the scattering length a ; \hbar is Planck's constant divided by 2π , m_e is the mass of the electron and δ the Dirac delta function. Given the Rydberg electron wavefunction $\Psi(r)$, the pseudopotential leads to a mean field potential:

$$V(R) = \frac{2\pi\hbar^2 a}{m_e} |\Psi(R)|^2 \quad (2)$$

The interaction is restricted to a range given by the size of the Rydberg atom, which is proportional to $n^2 a_0$, where a_0 denotes the Bohr radius. If this size is comparable to the mean interparticle distance, the interaction can lead to a bound state¹⁶ in the case of negative scattering lengths a . Diatomic¹⁷ and triatomic molecules¹⁸ of this class have been experimentally observed for ^{87}Rb atoms at principal quantum numbers around $n = 40$ and densities of the order of 10^{12} cm^{-3} . At higher densities, however, the lifetime of these bound states was found to decrease significantly¹⁹.

The experiments described here are performed in a completely different regime. We combine much higher Rydberg states at principal quantum numbers $n = 110$ – 202 with high densities of up to 10^{14} cm^{-3} in a BEC. In this regime, a significant fraction of the BEC is inside the electron orbit. Consequently, the single electron impurity has an impact on the BEC wavefunction as a whole. We vary the absolute number of atoms inside the electron orbit from 700 to up to 30,000 by changing the principal quantum number, which increases the spatial extent of the electron wavefunction up to the edges of the condensate (see Fig. 1).

We start with a magnetically trapped BEC of ^{87}Rb atoms in the $|5S_{1/2}, m_F = 2\rangle$ state. We illuminate the BEC with light from two lasers coupling the ground state to the $|nS_{1/2}, m_S = 1/2\rangle$ state ($n = 110$ – 202)

via the intermediate $5P_{3/2}$ state (see Methods). The strong van der Waals interaction between Rydberg atoms in such high states prevents the simultaneous excitation of two Rydberg atoms within a certain distance r_B . For Rydberg states with $n > 100$, this blockade effect²⁰ allows only a single Rydberg excitation in the whole BEC, because the blockade radius here is significantly larger than the Thomas–Fermi radii of the condensate (see Fig. 1a and Supplementary Information). This enables us to create one single electron impurity in the condensate. After a defined interaction time of the order of the Rydberg state lifetime, we apply an electric field pulse to remove the ionic core as well as the electron. We repeat this sequence several hundred times in a single condensate, before releasing the BEC and taking an absorption image after a time of flight of 50 ms. The long expansion time isolates the low-momentum components of the trapped atom cloud, as fast particles have already travelled out of the region of interest. From these time of flight images, we extract the change in atom number and aspect ratio of the BEC at different detunings of the excitation lasers relative to the single atom Rydberg level, and at different principal quantum numbers (see Fig. 2a, b).

We observe the impact of the electron impurity as a loss of atoms from the BEC after the time of flight. Compared to the Rydberg resonances measured in a thermal cloud at low densities the observed lines in the BEC are shifted to the red, with the magnitude of the shift being inversely proportional to the principal quantum number of the Rydberg atom. This shift is caused by the low-energy scattering of the single Rydberg electron from the BEC atoms. The contribution of the positively charged Rydberg core¹⁵ can be neglected in our case, because the range of this polarization potential is very small compared to the electron wavefunction (see Fig. 1b). Whereas the mean depth of the interaction

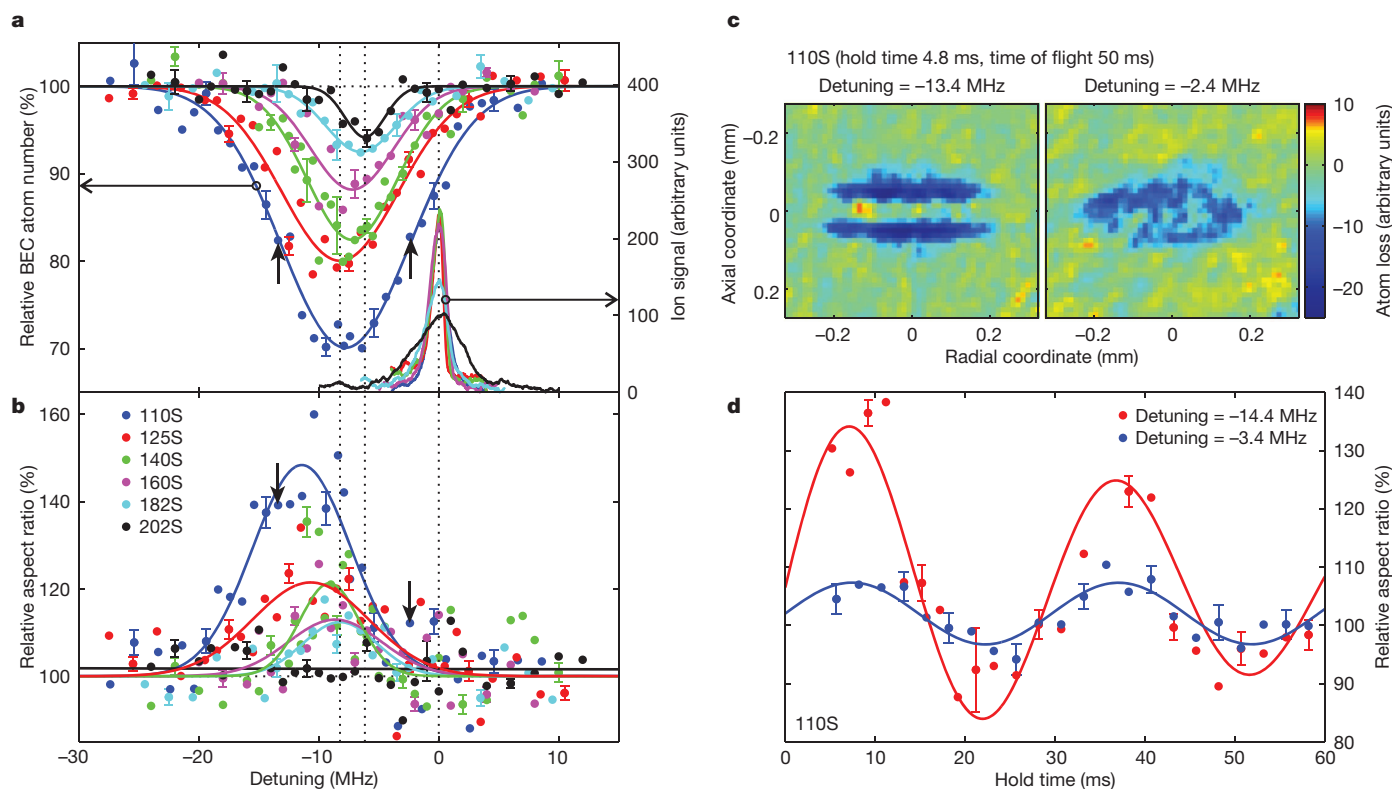


Figure 2 | Rydberg excitation spectra for different principal quantum numbers and the mechanical effect on the condensate. The relative BEC atom number (a) and aspect ratio (b) after the time of flight is plotted versus the laser detuning from the non-interacting Rydberg level, as determined from the ion signal in a low-density thermal sample. The spectral positions of the highest and lowest Rydberg state in the BEC and the zero position are indicated as vertical dotted lines. c, Images showing the atom loss (colour scale, right) and deformation of the condensate at two distinct spectral positions (marked by black arrows in a) of the 110S state, with the overall losses the same. Owing to

the excitation of a collective mode, the condensate becomes more elongated, compensating for the atom losses in the centre. d, The shape oscillation becomes obvious from a plot of the dependency of the aspect ratio change on the hold time in the trap; as examples, we show data for the 110S state at two different detunings, marked by black arrows in b (see Supplementary Information). The sequence lengths of the measurements in a, b and c correspond to hold times of 4.8 ms (110S and 125S) and 8.0 ms (140S–202S) respectively. Error bars, ± 1 s.d. from ten independent measurements.

potential decreases from 12 kHz to 290 Hz with increasing principal quantum number, the number of atoms inside the electron wavefunction increases. Integrating the mean field potential $V(R)$ over all atoms inside the Rydberg electron wavefunction we obtain the total shift, which depends to lowest order^{14,15} only on the mean atomic density \bar{n} :

$$\Delta E = \int_{\text{Ryd}} d^3 R V(R) n(R) = \frac{2\pi\hbar^2 a}{m_e} \bar{n} \quad (3)$$

For homogeneous densities the shift is therefore independent of the actual Rydberg state. This is consistent with our measurements at smaller principal quantum numbers where the local density approximation is fulfilled. For higher principal quantum numbers, however, we find a significant deviation. When the spatial extent of the Rydberg atom becomes comparable to the radial size of the BEC (see Fig. 1a), the electron is exploring mainly the low-density regions of the condensate. We can account for this by averaging the Thomas–Fermi density distribution over the size of the Rydberg atom. We assume the Rydberg atom is being excited in the centre of the condensate and use the average value of the peak density $\bar{n} = 8.6 \times 10^{13} \text{ cm}^{-3}$ over one sequence. In Fig. 3a the calculated shift is plotted in comparison to the line shifts extracted from Fig. 2a. In addition to the values for a constant scattering length²¹ of $a = -16.1 a_0$, a calculation taking higher-order scattering theory into account is shown (see also Supplementary Information). This simple model without any fitting parameters agrees quantitatively with our data.

The density-dependent shift of the Rydberg line allows us to control the position of the electron impurity inside the BEC with a precision much better than our optical resolution simply by changing the frequency of the excitation lasers. This is confirmed by the measured deformation of the condensate (Fig. 2b, c). The interaction of the electron impurity with the BEC excites a quadrupole shape oscillation (Fig. 2d and Supplementary Information) that leads to a change in the aspect ratio after the time of flight. In particular, we observe the strongest deformation on the red side of the BEC loss feature, that is, in the region of largest density-induced shift. Here the electron impurity is localized at the centre of the condensate and strong deformation of the condensate can be expected, whereas on the blue side and at high principal quantum numbers the mechanical effect averages to zero for repeated excitations. Quantitative modelling of this effect is, however, more involved and will be the subject of future studies.

To determine the actual interaction time of an individual impurity with the BEC, we measure the lifetime of the Rydberg atoms (see Fig. 3b and Supplementary Information). From empirical scaling laws²², we would expect the Rydberg lifetimes τ to increase with principal quantum number as n^3 , here from 1.7 ms (at $n = 110$) to 10.8 ms (at $n = 202$). However, we observe a reduced lifetime of around 780 μs at densities around 10^{12} cm^{-3} in the thermal cloud for all Rydberg states investigated between $n = 110$ and 202. In the condensate, at peak densities just below 10^{14} cm^{-3} , we find the lifetime to be further reduced by about two orders of magnitude. In this high-density regime we observe a lifetime increasing with principal quantum number. These two observations suggest that there is a dominant decay mechanism which is mainly dependent on the density of the gas. Further discussion of this effect can be found in Supplementary Information.

We now turn to the quantitative analysis of the fraction of atoms removed from the condensate owing to the electron impurity. To fully model this process, we have to take into account the coherent properties of the condensate and its excitations. We use perturbation theory to calculate the number of excitations in the condensate. Expressing the interaction potential $V(r)$ from equation (2) in terms of Fourier components $\rho_q = \int d^3 r e^{-iqr} |\Psi(r)|^2$ of the electron density, the relevant part of the interaction reads (see Supplementary Information):

$$\hat{H}_{\text{int}} = \frac{2\pi\hbar^2 a}{m_e} \frac{\bar{n}}{\sqrt{N}} \sum_{q \neq 0} \rho_q (u_q - v_q) (\hat{b}_q^\dagger + \hat{b}_{-q}) \quad (4)$$

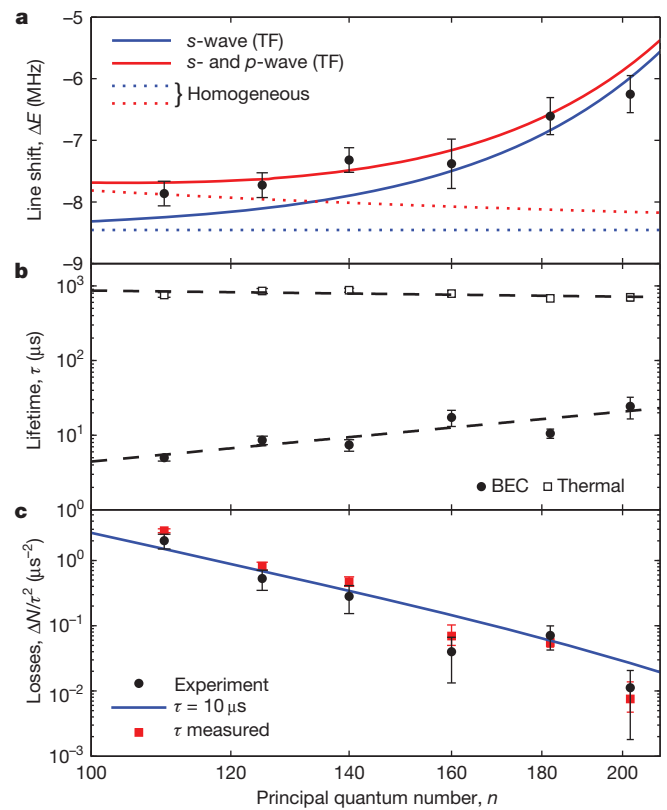


Figure 3 | Energy shift and lifetime reduction of the Rydberg state in a condensate and Rydberg electron induced loss of BEC atoms. **a**, Theory curves with constant *s*-wave scattering length (blue) and taking higher-order corrections into account (red). The corresponding values neglecting the Thomas–Fermi (TF) density distribution are indicated as dotted lines. **b**, Power-law fits (dashed lines) to lifetime measurements in the condensate (circles) and the thermal cloud (squares). **c**, Plots of the atom loss per pulse from Fig. 2a divided by the square of the Rydberg lifetime τ in the condensate. Theory values taking the measured lifetime into account (red squares) and assuming a constant lifetime of 10 μs (blue line) are shown. Error bars: 68% confidence bounds of Gaussian fits from Fig. 2a (a), exponential fits to the measured decay (b) and a combination of both (c).

where u_q and v_q denote the Bogoliubov factors and \hat{b}_q^\dagger is the creation operator of a collective excitation with quasi-momentum q . The finite lifetime τ of the Rydberg state leads to a time dependence of the perturbation, resulting in a finite Fourier width of the excitation. The shape of the potential $V(r)$ from equation (2) then determines which modes are actually populated within the allowed energy range. Because the outer edge of the electron density is located approximately at the Bohr orbit, which scales as $n^2 a_0$ (see Fig. 1b), the Fourier components ρ_q show a maximum at low momenta, clearly in the regime of phonon type excitations. However we find that the excitations at higher momenta within the Fourier width also lead to a significant contribution (see Supplementary Information). After the time of flight, both phonon and free-particle excitations can be detected as atom losses. Taking this conversion into account, we are able to reproduce our experimental results without any free parameters. Figure 3c shows the maximum atom loss extracted from the data in Fig. 2a, divided by the square of the lifetimes τ measured for each Rydberg state. This way we remove the main dependence on τ (see Supplementary Information). The solid line in Fig. 3c shows the atom loss predicted by our Bogoliubov calculation, assuming a constant lifetime of $\tau = 10 \mu\text{s}$ for all Rydberg states. This already reproduces the overall effect very well. Even better agreement between our data and theory can be achieved if we use the explicitly measured lifetime of each Rydberg state (red squares in Fig. 3c).

Our results on the coupling of a single electron to BEC excitations bring within reach the study of further phenomena, such as the trapping

of a whole condensate by an impurity. A repulsive interaction between electron and BEC could be achieved by changing the spin state of the electron²¹. Tiny electric fields are sufficient to deform and manipulate the Rydberg electron wavefunction, offering further ways to control the coupling. The interaction of the impurity with excitations already present in the condensate could provide a model system for phonon-mediated coupling of electrons. From the perspective of the single Rydberg atom, the strong interaction of the excited electron with the trapped atoms will influence its quantum mechanical state as well as its motional degrees of freedom. These effects enable intriguing quantum optics applications. For example, the scattering from the cold gas could serve as a source of dephasing, forming a crucial part of the proposal for a single photon absorber²³. Last, the BEC provides a sensitive probe for the detection of individual Rydberg atoms, because the depletion of the BEC is localized inside the volume of the electron wavefunction. Highly ordered quasi-crystalline samples of multiple Rydberg atoms in one condensate or even two overlapping Rydberg atoms could be created and probed by means of double resonance spectroscopy²⁴. By using well established techniques, such as *in situ* phase contrast imaging, even the imaging of a single electron orbital seems feasible.

METHODS SUMMARY

We start with a condensate of $N = 8 \times 10^4$ atoms of ^{87}Rb in the $|5S_{1/2}, m_F = 2\rangle$ state in a cloverleaf-type magnetic trap at a high magnetic offset field of 1.355 mT (radial and axial trap frequencies are respectively $\omega_\rho = 2\pi \times 81.7$ Hz and $\omega_z = 2\pi \times 22.4$ Hz). We excite Rydberg states $|nS_{1/2}, m_S = 1/2\rangle$ (with $n = 110\text{--}202$) via a two-photon transition detuned by 500 MHz from the intermediate $5P_{3/2}$ state using continuous wave diode lasers at wavelengths of 780 nm and 480 nm. The blue laser beam has a power of 100 mW and is focused down to a size of 120 μm ($1/e^2$ diameter). We choose the power of the red beam to be typically in the range of 3 μW at a $1/e^2$ diameter of 1 mm. The two laser beams are counter-propagating along the magnetic field axis of the trap. We address the desired transition by choosing σ^+ and σ^- polarization for the 780-nm and 480-nm laser, respectively.

In each condensate, a sequence consisting of a 1- μs light pulse for Rydberg excitation and a 2- μs electric field pulse for removal of any remaining Rydberg atoms or ions, separated by 10- μs delay time, is repeated 200 (Fig. 2d), 300 (110S and 125S states) or 500 (other states) times at a rate of 62.5 kHz. The clearance field is set to 5.7 V cm^{-1} , well above the ionization threshold of all Rydberg states under investigation. The magnetic trap is switched off immediately after the Rydberg sequence. After a time of flight of 50 ms we take an absorption image of the condensate. We determine the relative change in atom number and aspect ratio by comparing each measurement with a consecutive reference image in which the blue Rydberg laser is detuned by more than 40 MHz.

Further details about the set-up and the data analysis can be found in ref. 25 and in Supplementary Information.

Received 6 June; accepted 20 August 2013.

1. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
2. Reif, F. & Meyer, L. Study of superfluidity in liquid He by ion motion. *Phys. Rev.* **119**, 1164–1173 (1960).
3. Rayfield, G. W. & Reif, F. Evidence for the creation and motion of quantized vortex rings in superfluid helium. *Phys. Rev. Lett.* **11**, 305–308 (1963).

4. Fisher, D. S., Halperin, B. I. & Platzman, P. M. Phonon-ripplon coupling and the two-dimensional electron solid on a liquid-helium surface. *Phys. Rev. Lett.* **42**, 798–801 (1979).
5. Platzman, P. M. & Dykman, M. I. Quantum computing with electrons floating on liquid helium. *Science* **284**, 1967–1969 (1999).
6. Robert, A. *et al.* A Bose-Einstein condensate of metastable atoms. *Science* **292**, 461–464 (2001).
7. Ciampini, D. *et al.* Photoionization of ultracold and Bose-Einstein-condensed Rb atoms. *Phys. Rev. A* **66**, 043409 (2002).
8. Zipkes, C., Palzer, S., Sias, C. & Köhl, M. A trapped single ion inside a Bose-Einstein condensate. *Nature* **464**, 388–391 (2010).
9. Ratschbacher, L., Zipkes, C., Sias, C. & Köhl, M. Controlling chemical reactions of a single particle. *Nature Phys.* **8**, 649–652 (2012).
10. Härter, A. *et al.* Single ion as a three-body reaction center in an ultracold atomic gas. *Phys. Rev. Lett.* **109**, 123201 (2012).
11. Shin, Y., Schunck, C. H., Schirotzek, A. & Ketterle, W. Tomographic rf spectroscopy of a trapped Fermi gas at unitarity. *Phys. Rev. Lett.* **99**, 090403 (2007).
12. Schirotzek, A., Wu, C.-H., Sommer, A. & Zwierlein, M. W. Observation of Fermi polarons in a tunable Fermi liquid of ultracold atoms. *Phys. Rev. Lett.* **102**, 230402 (2009).
13. Massignan, P., Pethick, C. J. & Smith, H. Static properties of positive ions in atomic Bose-Einstein condensates. *Phys. Rev. A* **71**, 023606 (2005).
14. Amaldi, E. & Segrè, E. Effect of pressure on high terms of alkaline spectra. *Nature* **133**, 141 (1934).
15. Fermi, E. Sopra lo spostamento per pressione delle righe elevate delle serie spettrali. *Nuovo Cim.* **11**, 157–166 (1934).
16. Greene, C. H., Dickinson, A. S. & Sadeghpour, H. R. Creation of polar and nonpolar ultralong-range Rydberg molecules. *Phys. Rev. Lett.* **85**, 2458–2461 (2000).
17. Bendkowsky, V. *et al.* Observation of ultralong-range Rydberg molecules. *Nature* **458**, 1005–1008 (2009).
18. Bendkowsky, V. *et al.* Rydberg trimers and excited dimers bound by internal quantum reflection. *Phys. Rev. Lett.* **105**, 163201 (2010).
19. Butscher, B. *et al.* Lifetimes of ultralong-range Rydberg molecules in vibrational ground and excited states. *J. Phys. At. Mol. Opt. Phys.* **44**, 184004 (2011).
20. Saffman, M., Walker, T. G. & Mølmer, K. Quantum information with Rydberg atoms. *Rev. Mod. Phys.* **82**, 2313–2363 (2010).
21. Bahrim, C., Thumm, U. & Fabrikant, I. I. $^3\text{S}_0$ and $^1\text{S}_0$ scattering lengths for e^- + Rb, Cs and Fr collisions. *J. Phys. At. Mol. Opt. Phys.* **34**, L195–L201 (2001).
22. Beterov, I. I., Ryabtsev, I. I., Tretyakov, D. B. & Entin, V. M. Quasiclassical calculations of blackbody-radiation-induced depopulation rates and effective lifetimes of Rydberg ns , np , and nd alkali-metal atoms with $n \leq 80$. *Phys. Rev. A* **79**, 052504 (2009).
23. Honer, J., Löw, R., Weimer, H., Pfau, T. & Büchler, H. P. Artificial atoms can do more than atoms: deterministic single photon subtraction from arbitrary light fields. *Phys. Rev. Lett.* **107**, 093601 (2011).
24. Reinhard, A. *et al.* Double-resonance spectroscopy of interacting Rydberg-atom systems. *Phys. Rev. Lett.* **100**, 233201 (2008).
25. Löw, R. *et al.* An experimental and theoretical guide to strongly interacting Rydberg gases. *J. Phys. At. Mol. Opt. Phys.* **45**, 113001 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Rzążewski and J. Hecker Denschlag for discussions and C. Tresp for setting up the Rydberg laser system. This work was funded by the Deutsche Forschungsgemeinschaft (DFG) within SFB/TRR21 and project PF 381/4-2. We also acknowledge support by the ERC under contract number 267100, and A.G. acknowledges support from EU Marie Curie programme ITN-Coherence 265031. S.H. is supported by the DFG through project HO 4787/1-1.

Author Contributions The experiment was conceived by J.B.B., R.L., S.H. and T.P. and carried out by J.B.B., A.T.K. and A.G.; data analysis was accomplished by J.B.B., A.T.K. and A.G.; perturbation theory was developed by D.P. and H.P.B.; and J.B.B. and D.P. wrote the manuscript with contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.P. (t.pfau@physik.uni-stuttgart.de).

Gradual demise of a thin southern Laurentide ice sheet recorded by Mississippi drainage

Andrew D. Wickert¹, Jerry X. Mitrovica², Carlie Williams³ & Robert S. Anderson¹

At the Last Glacial Maximum (LGM), about 21,000 years before present, land-based ice sheets held enough water to reduce global mean sea level by 130 metres¹. Yet after decades of study, major uncertainties remain as to the distribution of that ice². Here we test four reconstructions of North American deglacial ice-sheet history^{3–6} by quantitatively connecting them to high-resolution oxygen isotope ($\delta^{18}\text{O}$) records from the Gulf of Mexico^{7–11} using a water mixing model¹². For each reconstruction, we route meltwater^{3–6} and seasonal runoff^{13–16} through the time-evolving Mississippi drainage basin, which co-evolves with ice geometry^{3–6} and changing topography as ice loads deform the solid Earth and produce spatially variable sea level in a process known as glacial isostatic adjustment¹⁷. The $\delta^{18}\text{O}$ records show that the Mississippi-drained southern Laurentide ice sheet contributed only 5.4 ± 2.1 metres to global sea level rise, of which 0.66 ± 0.07 metres were released during the meltwater pulse 1A event 14,650–14,310 years before present¹⁸, far less water than previously thought^{5,12,19}. In contrast, the three reconstructions based on glacial isostatic adjustment^{3–5} overpredict the $\delta^{18}\text{O}$ -based post-LGM meltwater volume by a factor of 1.6 to 3.6. The fourth reconstruction⁶, which is based on ice physics, has a low

enough Mississippi-routed meltwater discharge to be consistent with $\delta^{18}\text{O}$ constraints, but also contains the largest LGM North American ice volume. This suggests that modelling based on ice physics may be the best way of matching isotopic records while also sequestering enough water in the North American ice sheets to match the observed LGM sea level fall¹.

Existing methods for reconstructing ice-sheet histories since the Last Glacial Maximum (LGM) can produce different ice sheet thickness patterns that match presently-used observational constraints. This non-uniqueness in ice sheet reconstructions presents a major problem in understanding global deglaciation and the associated 130 m (ref. 1) of global mean sea level rise since the LGM. Geophysical inferences based on sea level modelling^{4,5} are limited by both the spatial and temporal scarcity of available relative-sea-level histories and uncertainties in the viscoelastic structure of the Earth's mantle²⁰. Moreover, ice-physics-based models do not yet include realistic treatments of ice-stream dynamics, and they are sensitive to uncertainties in climate forcing, basal conditions and ice-calving dynamics⁶. Although combining these two approaches may reduce their respective potentials for geophysical and glaciological inconsistency^{21,22}, fundamental sources of non-uniqueness

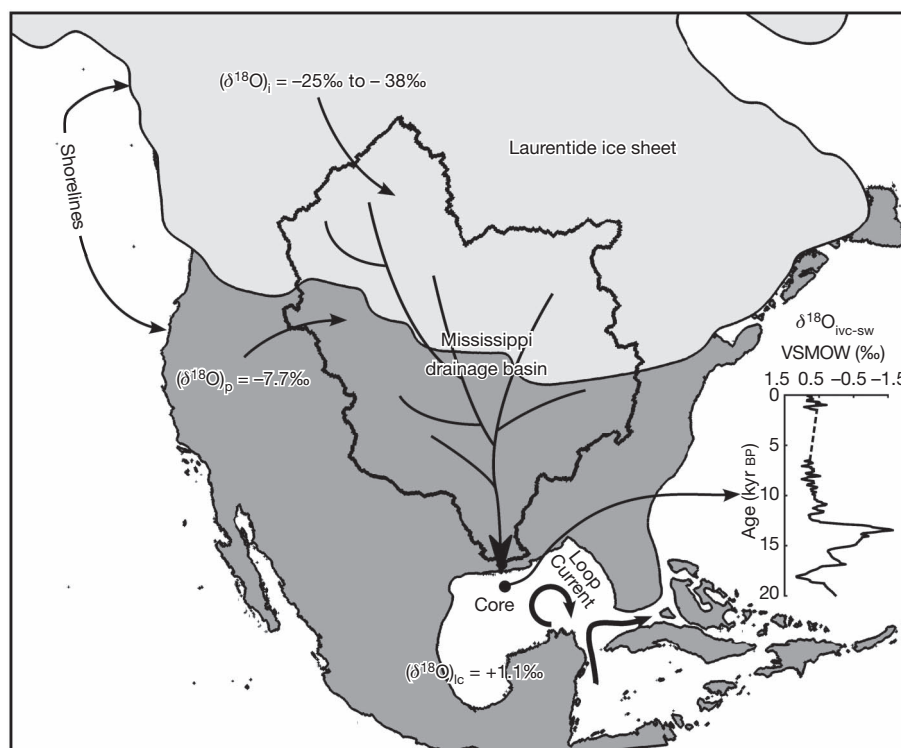


Figure 1 | Construction of the oxygen isotope record. Gulf of Mexico waters are a mix (equation (1)) of fluvial inputs—seasonal precipitation ($\delta^{18}\text{O}_p = -7.7\text{‰}$) and ice-sheet melt ($\delta^{18}\text{O}_i = -25\text{‰}$ (ref. 29) to -38‰

(ref. 30))—and of water advected into the Gulf by eddies shed by the loop current ($\delta^{18}\text{O}_e = 1.1\text{‰}$). The ice-sheet extent and shorelines are based on ICE-5G/VM2 (ref. 5) at 14.5 kyr BP. The inset shows the moving average $\delta^{18}\text{O}_{\text{IVC-SW}}^{7-11}$ (see text).

¹Institute of Arctic and Alpine Research and Department of Geological Sciences, University of Colorado, 1560 30th Street, Boulder, Colorado 80303, USA. ²Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA. ³College of Marine Science, University of South Florida, 140 7th Avenue South, St Petersburg, Florida 33701, USA.

remain. In this study, we describe a general method of testing ice-sheet reconstructions using a sequence of self-consistent physical models to compare meltwater discharge histories derived from ice reconstructions against chemostratigraphic records from river outlets.

As a first application, we test four published models of the evolution of the Laurentide ice sheet (LIS)^{3–6} against $\delta^{18}\text{O}$ measurements from marine sediments in the Gulf of Mexico^{7–11} that record LIS discharge routed down the Mississippi River. We pay particular attention to meltwater pulse 1A (MWP-1A), because the dominant source for this 14.65–14.31-kyr-BP episode of catastrophic ice-sheet collapse, which produced 14–18 m of global mean sea level rise (about 47 mm yr^{-1})¹⁸, remains a matter of significant contention^{2,12,18,22,23}.

Ice-age river systems are the threads that tie ice sheets to the oceans and marine sedimentary records. These riverine threads are stretched, severed and reconnected by evolving ice-sheet geometry and geographically variable post-glacial sea level change. It is necessary to address this coupled system to reconstruct drainage accurately, especially in the flat-lying North American cratonal interior where regional slopes are dominated by ice sheets and glacial isostatic adjustment (GIA). We route flow over a 30-arcsecond digital elevation model of the Earth's surface topography and bathymetry that is summed with models of ice-sheet thickness^{3–6} and deformed by these models' associated predictions of sea level change and GIA (see Methods)¹⁷. The 30-arcsecond resolution is essential for accurate drainage reconstructions because the ~ 600 – 900 -m-wide pixels are required to resolve river valleys. The ice models include ICE-3G³, ICE-5G⁵, a model developed at the Australian National University⁴ ('ANU') and a model of the Cordilleran–Laurentide–Greenland ice-sheet system derived from dynamical ice-sheet modelling⁶ ('G12'). Meteoric water inputs (precipitation minus evapotranspiration) to the drainage history are obtained from the model output of the TraCE-21K continuous general-circulation-model run^{14,15}, calibrated to modern measurements^{13,16} and the modern mean Mississippi discharge of $16,790 \text{ m}^3 \text{ s}^{-1}$.

We combine these time-variable runoff histories (Extended Data Fig. 1) with ocean water from the Gulf of Mexico using a mixing model¹² (Fig. 1) to forward-calculate synthetic $\delta^{18}\text{O}$ histories for each of the ice reconstructions (Fig. 2a). The Gulf of Mexico Loop Current input discharge (Fig. 1) is assumed to be constant with time, and its value is constrained such that the modern Mississippi discharge is consistent with near-modern $\delta^{18}\text{O}$ data⁹.

Figure 3 shows a dynamic Mississippi River drainage basin that approaches twice its current size at the LGM, and shrinks and swells in response to the evolving ice-sheet geometry and GIA. All four drainage reconstructions demonstrate that the modern Mississippi drainage basin contains the downstream remnants of many once-larger river systems that dramatically evolved in size and discharge during glacial–interglacial cycles. Different models of deglaciation and GIA produce a wide range of modelled drainage basin geometries during MWP-1A and the onset of the Younger Dryas cold period, illustrating how subtle changes in the tilt of the Great Lakes and mid-continent can drastically affect patterns of continental-scale runoff.

To assess the veracity of these simulations, we have compiled $\delta^{18}\text{O}$ records from the Gulf of Mexico^{7–11}. High-temporal-resolution $\delta^{18}\text{O}$ measurements with tight age controls are necessary to prevent infrequent and short-lived high-discharge events from being misinterpreted as extended periods of rapid ice-sheet melt. Furthermore, the foraminiferal $\delta^{18}\text{O}$ data must be corrected for both sea surface temperature, to isolate the $\delta^{18}\text{O}$ of sea water, and global ice volume, to account for the $\sim 1\text{‰}$ change in sea-water $\delta^{18}\text{O}$ between the LGM and the present due to the preferential accumulation of the light ^{16}O isotope in continental ice. The resulting temperature- and ice-volume-corrected sea-water $\delta^{18}\text{O}$ data ($\delta^{18}\text{O}_{\text{IVC-SW}}$) from the Gulf of Mexico are primarily controlled by Mississippi River discharge. Our compilation (Fig. 2a) includes only those data that meet all of these criteria^{7–11}.

The $\delta^{18}\text{O}_{\text{IVC-SW}}$ record shows evidence of ice-sheet ablation in the Mississippi drainage basin about 20 kyr before present (BP), followed by a short-lived period from 18.6 kyr BP to 17.5 kyr BP of $1.6 \pm 0.4 \text{ m}$ sea level

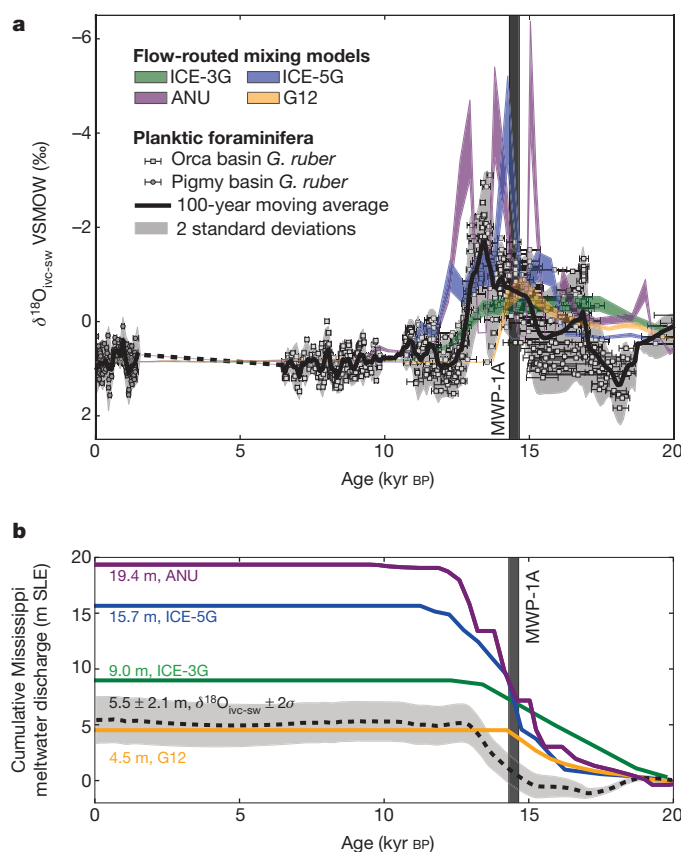


Figure 2 | Data–model intercomparison. **a**, River discharges from ice-sheet reconstructions^{3–6} converted into $\delta^{18}\text{O}_{\text{IVC-SW}}$ for comparison with data from refs 7–11. The spread in $\delta^{18}\text{O}_{\text{IVC-SW}}$ for each ice-sheet reconstruction (shaded areas) reflects the -25‰ to -38‰ bounds on LIS $\delta^{18}\text{O}$ (refs 29, 30). $\delta^{18}\text{O}_{\text{IVC-SW}}$ data were binned into 100-year windows using a Gaussian weighting function based on the means and standard deviations of sample ages; error bars are 2σ . The data show a period of southern LIS growth around 18.0 kyr BP, and spikes of meltwater release at about 16.8 kyr BP and 13.4 kyr BP, none of which is reproduced in any of the models. G12 follows the observed trend of increasing meltwater discharge that is rapidly rerouted away from the Mississippi, but this rerouting occurs about 1.1 kyr earlier than in the geologic record. Through the Holocene, both the data and models show consistent near-modern Mississippi River discharge. **b**, Modelled and $\delta^{18}\text{O}_{\text{IVC-SW}}$ -based cumulative Mississippi meltwater discharge contributions to sea level rise (shown in metres SLE).

equivalent (SLE) ice-sheet growth that correlates with the Port Bruce Readvance of the southern LIS into the Great Lakes basins²⁴ but does not appear in any of the ice models we test^{3–6}. (Ice volumes are back-calculated from $\delta^{18}\text{O}_{\text{IVC-SW}}$ using equation (1) and presented as equivalent liquid water depth across the world ocean (SLE).) A short-lived meltwater peak at 16.8 kyr BP interrupted the cold 'Mystery Interval' stadial (17.5–14.5 kyr BP)¹¹, and coincides with both flooding from Lake Superior²⁵ and southwestern LIS reorganization into a thin ice-stream system that produced rapidly readjusting ice lobes²⁶. $\delta^{18}\text{O}_{\text{IVC-SW}}$ values increased from 16.8 kyr BP to 15.4 kyr BP, indicating decreasing meltwater discharge down the Mississippi; these dates encompass the timing of the eastward rerouting of Great Lakes drainage during the Mackinaw interstade²⁴. At 15.4 kyr BP, meltwater discharge rose strongly, and remained high through the warm Bølling–Allerød interstadial (about 14.5–12.9 kyr BP)¹¹. It peaked from 13.7 to 12.9 kyr BP, coincident with outflow from lakes Agassiz, Superior and Michigan into the Mississippi^{24,27}. None of the ice models produced discharge peaks during this time, perhaps owing to their non-inclusion of proglacial lakes and relevant ice-calving dynamics. Meltwater discharge then fell abruptly at the onset of the Younger Dryas (about 12.9 kyr BP¹¹) owing to widespread rerouting of LIS meltwater away from the Mississippi^{11,24,27}.

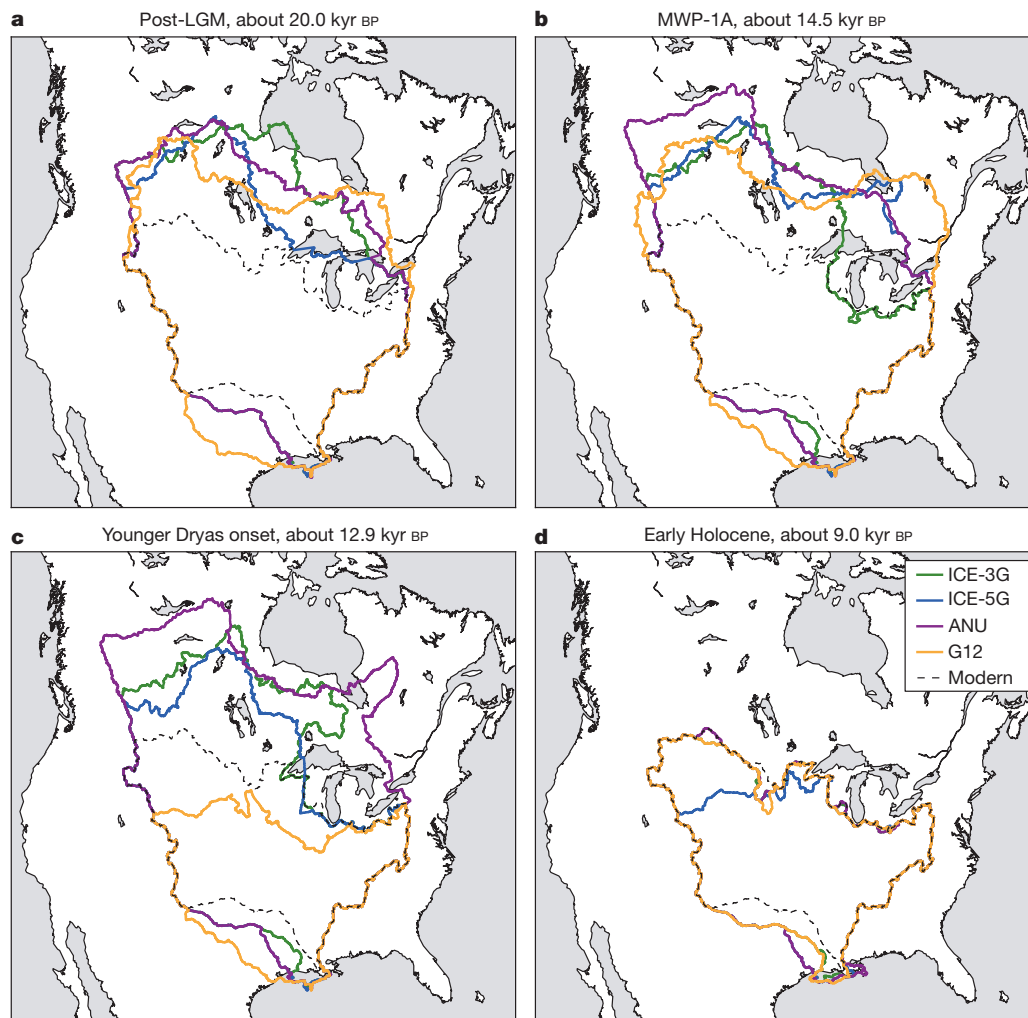


Figure 3 | Reconstructed evolution of the Mississippi drainage basin.

Drainage patterns evolve as a function of melting ice sheets^{3–6} and the glacial isostatic adjustment (GIA) process¹⁷. Differences between the ice models and their respective GIA responses produce major differences in retrodicted Mississippi drainage basin extent, especially across the flat-lying North American midcontinent where large proglacial lakes strongly affected ice-sheet retreat and meltwater discharge^{24,27}. **a**, Mississippi drainage basin extent after the Last Glacial Maximum for the following models and their times. ICE-3G,

20.24 kyr BP. ICE-5G, 20.0 kyr BP. ANU, 20.0 kyr BP. G12, 20.0 kyr BP.

b, Mississippi drainage basin extent during MWP-1A for the following models and their times. ICE-3G, 13.8 kyr BP. ICE-5G, 14.5 kyr BP. ANU, 14.38 kyr BP. G12, 14.5 kyr BP. **c**, Mississippi drainage basin extent at the onset of the Younger Dryas for the following models and their times. ICE-3G, 13.01 kyr BP. ICE-5G, 13.0 kyr BP. ANU, 12.8 kyr BP. G12, 13.0 kyr BP. **d**, Mississippi drainage basin extent during the Early Holocene for the following models and their times. ICE-3G, 8.81 kyr BP. ICE-5G, 9.0 kyr BP. ANU, 9.05 kyr BP. G12, 9.0 kyr BP.

We turn to the central goal of the study: an assessment of published models for the LIS based on comparisons with the observed $\delta^{18}\text{O}_{\text{IVC-SW}}$ record. The $\delta^{18}\text{O}_{\text{IVC-SW}}$ record (Fig. 2a) implies a net Mississippi meltwater discharge of 5.4 ± 2.1 m SLE (2σ error; Fig. 2b). The ICE-3G, ICE-5G and ANU models respectively overpredict this discharge to be 9.0 m, 15.7 m and 19.4 m SLE. The total Mississippi-routed meltwater output of the G12 model—4.5 m SLE—falls within the error of the $\delta^{18}\text{O}_{\text{IVC-SW}}$ -based prediction (Fig. 2b, Extended Data Table 1), making this ice-physics-based model the only one to fit the data.

Our work adds to the constraints on MWP-1A. The $\delta^{18}\text{O}_{\text{IVC-SW}}$ data predict that 0.66 ± 0.07 m SLE of meltwater were routed down the Mississippi during the 340-year-long¹⁸ MWP-1A time window (Fig. 2b), a small fraction of the approximately 14–18 m SLE released during this period. We exclude potential hyperpycnal water inputs^{12,19,21} because several lines of evidence suggest that they are insignificant (see Methods). The observed steady, mildly elevated meltwater discharge through MWP-1A (Fig. 2b) is consistent with geological evidence of a thin and actively streaming southern LIS at this time^{24,26}, and corroborates sea-level fingerprinting that rules out the southern LIS as the sole source for MWP-1A²³.

However, other ice sheets do not seem to be likely sources for MWP-1A. New modelling of GIA in Antarctica suggests that its ice sheet contained an excess LGM volume of no more than ~ 8 m SLE²⁸, limiting its potential contribution to MWP-1A. Likewise, the Eurasian ice sheets probably contributed only about 4.1–5.7 m SLE² to MWP-1A. If these estimates are robust, other sectors of the LIS must be major contributors to MWP-1A.

Of the ice models in our study, ICE-3G does not include MWP-1A, and therefore cannot match global sea level data⁴. ICE-5G and ANU predict 3.30 m and 1.73 m SLE, respectively, discharged down the Mississippi during MWP-1A, significantly overpredicting the $\delta^{18}\text{O}_{\text{IVC-SW}}$ record (Fig. 2b, Extended Data Table 1). G12 predicts a Mississippi-routed MWP-1A contribution of 0.61 m SLE, which is within the error of the observations.

The ice-physics-driven G12 ice sheet has the largest North American LGM ice volume (Extended Data Table 1). It can therefore offset a smaller Antarctic excess ice volume²⁸. At the same time, it routes less meltwater through the Mississippi, matching both the total integrated and MWP-1A-specific $\delta^{18}\text{O}_{\text{IVC-SW}}$ -derived meltwater histories^{7–11}. G12 produces a rapid eastward shift of deglacial drainage away from the Mississippi,

consistent with geologic observations of rapid abandonment of the southern outlet of the glacial lake Agassiz^{11,27}, but places it at about 14 kyr BP, which is 1.1 kyr before it appears in the geologic record at the onset of the Younger Dryas^{11,27}. Although the authors of G12 advocate shifting its timing of deglaciation back by 2.5 kyr to associate ice-sheet saddle collapse with MWP-1A⁶ (see Methods), this would increase its misfit to the observed drainage rerouting¹¹. Nevertheless, the general success of (the unshifted) G12 in matching a number of constraints shows that ice-physics-based models^{6,21,22} hold great promise for reconstructions of past ice sheets.

Our general method of tracking the evolution of drainage basins and inferring meltwater discharge through specific outlets complements GIA-based reconstructions of ancient ice cover. The GIA process produces changes in topography that are filtered by lithospheric flexure and are sensitive to both mantle rheology and ice-sheet thickness²⁰. Surface-water drainage, on the other hand, operates within discrete drainage basin boundaries (Fig. 3) that are controlled primarily by ice-sheet geometry and leave a measurable stratigraphic record at their outlets. Our use of the oxygen isotope record in the Gulf of Mexico, associated with Mississippi River discharge, highlights significant deficiencies in previously published reconstructions of North American ice cover since the LGM, but also provides a guide for how newer models of this ice cover⁶ must be revised to reconcile the chemostratigraphic, sea-level, glaciological and climatic records of the last deglaciation event of the Pleistocene ice age.

METHODS SUMMARY

We connected ice-sheet-reconstruction-based retrodictions of meltwater discharge to the Gulf of Mexico, whose oxygen isotope record provides an additional constraint on past ice-sheet geometries. First, we generated palaeotopographies by combining modern surface elevations at 30-arcsecond resolution, ice-sheet thicknesses^{3–6} interpolated to 30-arcsecond resolution, and GIA¹⁷. We then reconstructed a time series of Mississippi drainage basin extents. At each time step, we computed the Mississippi discharge from seasonal precipitation minus evapotranspiration by combining the TraCE-21K LGM-to-present continuous run of the Community Climate System Model version 3^{14,15} with present-day measurements^{13,16}, and constrained the results to match the modern mean Mississippi discharge of $16,790 \text{ m}^3 \text{ s}^{-1}$ at the 0-kyr-BP time step. Mississippi discharge from ice-sheet ablation was added by differentiating ice-sheet volume between adjacent ice-model^{3–6} time steps. Both precipitation minus evapotranspiration and meltwater inputs were integrated across the time-evolving Mississippi drainage basin.

To test these simulated discharges, we compiled available $\delta^{18}\text{O}$ data from the Gulf of Mexico^{7–11}. We corrected these data for sea surface temperatures^{8–10} and global ice volume⁷ to isolate river inputs¹¹. We termed these ice-volume-corrected sea-water data ' $\delta^{18}\text{O}_{\text{IVC-SW}}$ '.

We adopted a water and oxygen isotope mixing model¹² to link ice-sheet-reconstruction-derived river discharges and the $\delta^{18}\text{O}_{\text{IVC-SW}}$ data. Of the water inputs, Q_i is river discharge from ice-sheet melt (-25% (ref. 29) to -38% (ref. 30)), Q_p is river discharge from seasonal precipitation minus evapotranspiration (-7.7% ; ref. 19), and Q_{lc} is ocean water input from the Gulf of Mexico Loop Current ($+1.1\%$; ref. 19).

$$\delta^{18}\text{O}_{\text{IVC-SW}} = \frac{(\delta^{18}\text{O})_i Q_i + (\delta^{18}\text{O})_p Q_p + (\delta^{18}\text{O})_{lc} Q_{lc}}{Q_i + Q_p + Q_{lc}} \quad (1)$$

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 June; accepted 28 August 2013.

1. Austermann, J., Mitrovica, J. X., Latychev, K. & Milne, G. A. Barbados-based estimate of ice volume at Last Glacial Maximum affected by subducted plate. *Nature Geosci.* **6**, 553–557 (2013).
2. Carlson, A. E. and Clark, P. U. Ice sheet sources of sea level rise and freshwater discharge during the last deglaciation. *Rev. Geophys.* **50**, RG4007 (2012).
3. Tushingham, A. M. & Peltier, W. R. Ice-3G: a new global model of Late Pleistocene deglaciation based upon geophysical predictions of post-glacial relative sea level change. *J. Geophys. Res.* **96**, 4497–4523 (1991).
4. Lambeck, K., Yokoyama, Y. & Purcell, T. Into and out of the Last Glacial Maximum: sea-level change during Oxygen Isotope Stages 3 and 2. *Quat. Sci. Rev.* **21**, 343–360 (2002).

5. Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32**, 111–149 (2004).
6. Gregoire, L. J., Payne, A. J. & Valdes, P. J. Deglacial rapid sea level rises caused by ice-sheet saddle collapses. *Nature* **487**, 219–222 (2012).
7. Flower, B. P., Hastings, D. W., Hill, H. W. & Quinn, T. M. Phasing of deglacial warming and Laurentide Ice Sheet meltwater in the Gulf of Mexico. *Geology* **32**, 597 (2004).
8. LoDico, J. M., Flower, B. P. & Quinn, T. M. Subcentennial-scale climatic and hydrologic variability in the Gulf of Mexico during the early Holocene. *Paleoceanography* **21**, PA3015 (2006).
9. Richey, J., Poore, R., Flower, B. & Quinn, T. 1400 yr multiproxy record of climate variability from the northern Gulf of Mexico. *Geology* **35**, 423–426 (2007).
10. Williams, C., Flower, B. & Hastings, D. Deglacial abrupt climate change in the Atlantic Warm Pool: a Gulf of Mexico perspective. *Paleoceanography* **25**, PA4221 (2010).
11. Williams, C., Flower, B. P. & Hastings, D. W. Seasonal Laurentide Ice Sheet melting during the “Mystery Interval” (17.5–14.5 ka). *Geology* **40**, 955–958 (2012).
12. Carlson, A. E. Geochemical constraints on the Laurentide Ice Sheet contribution to Meltwater Pulse 1A. *Quat. Sci. Rev.* **28**, 1625–1630 (2009).
13. Xie, P. & Arkin, P. A. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Am. Meteorol. Soc.* **78**, 2539–2558 (1997).
14. Liu, Z. *et al.* Transient simulation of last deglaciation with a new mechanism for Bolling-Allerod warming. *Science* **325**, 310–314 (2009).
15. He, F. *Simulating Transient Climate Evolution of the Last Deglaciation with CCSM3*. PhD thesis, Univ. Wisconsin (2010).
16. Mu, Q., Zhao, M. & Running, S. W. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* **115**, 1781–1800 (2011).
17. Kendall, R. A., Mitrovica, J. X. & Milne, G. A. On post-glacial sea level-II. Numerical formulation and comparative results on spherically symmetric models. *Geophys. J. Int.* **161**, 679–706 (2005).
18. Deschamps, P. *et al.* Ice-sheet collapse and sea-level rise at the Bolling warming 14,600 years ago. *Nature* **483**, 559–564 (2012).
19. Aharon, P. Entrainment of meltwaters in hyperpycnal flows during deglaciation superloods in the Gulf of Mexico. *Earth Planet. Sci. Lett.* **241**, 260–270 (2006).
20. Mitrovica, J. X. Haskell [1935] revisited. *J. Geophys. Res.* **101**, 555–569 (1996).
21. Tarasov, L. & Peltier, W. R. A calibrated deglacial drainage chronology for the North American continent: evidence of an Arctic trigger for the Younger Dryas. *Quat. Sci. Rev.* **25**, 659–688 (2006).
22. Tarasov, L., Dyke, A. S., Neal, R. M. & Peltier, W. A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling. *Earth Planet. Sci. Lett.* **315–316**, 30–40 (2012).
23. Clark, P., Mitrovica, J., Milne, G. & Tamisiea, M. Sea-level fingerprinting as a direct test for the source of global meltwater pulse 1A. *Science* **295**, 2438–2441 (2002).
24. Liciardi, J. M., Teller, J. T. & Clark, P. U. in *Mechanisms of Global Climate Change at Millennial Time Scales* (eds Clark, U., Webb, S. & Keigwin, D.) Vol. 112, 177–201 (Geophys. Monogr. 12, AGU, 1999).
25. Knox, J. C. in *Large Rivers: Geomorphology and Management* (ed. Gupta, A.) 145–182 (Wiley, 2007).
26. Ross, M., Campbell, J. E., Parent, M. & Adams, R. S. Palaeo-ice streams and the subglacial landscape mosaic of the North American mid-continental prairies. *Boreas* **38**, 421–439 (2009).
27. Teller, J. T. & Leverington, D. W. Glacial Lake Agassiz: a 5000 yr history of change and its relationship to the $\delta^{18}\text{O}$ record of Greenland. *Geol. Soc. Am. Bull.* **116**, 729–742 (2004).
28. Whitehouse, P. L., Bentley, M. J. & Le Brocq, A. M. A deglacial model for Antarctica: geological constraints and glaciological modelling as a basis for a new model of Antarctic glacial isostatic adjustment. *Quat. Sci. Rev.* **32**, 1–24 (2012).
29. Remenda, V. H., Cherry, J. A. & Edwards, T. W. D. Isotopic composition of old ground water from Lake Agassiz: implications for Late Pleistocene climate. *Science* **266**, 1975–1978 (1994).
30. Hooke, R. L. & Clausen, H. B. Wisconsin and Holocene $\delta^{18}\text{O}$ variations, Barnes Ice Cap, Canada. *Geol. Soc. Am. Bull.* **93**, 784–789 (1982).

Acknowledgements We thank F. He, B. Otto-Bliesner and Z. Liu for supplying their TraCE-21K general circulation model outputs. The Climate Prediction Center Merged Analysis of Precipitation (CMAP) precipitation data were provided by the NOAA/OAR/ESRL PSD from their website at <http://www.esrl.noaa.gov/psd/>. A.D.W. was supported by the US Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program, and by the US National Science Foundation Graduate Research Fellowship under grant number DGE 1144083. J.X.M. acknowledges support from the Canadian Institute for Advanced Research and Harvard University.

Author Contributions A.D.W. built and ran the drainage basin analysis routine, compiled and corrected the $\delta^{18}\text{O}$ data, performed the data-model comparisons, and interpreted the results. J.X.M. provided the global sea level model outputs and post-processing software. C.W. produced a large part of the $\delta^{18}\text{O}$ data. R.S.A. assisted with idea development. A.D.W. wrote the manuscript, with input and suggestions from R.S.A., J.X.M. and C.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D.W. (wickert@colorado.edu).

METHODS

The analysis in the main text consists of (1) retrodicting palaeotopographies and drainage basin areas over the post-LGM window; (2) calculating Mississippi River discharge due to meltwater and seasonal precipitation minus evapotranspiration at each time step; and (3) combining this drainage history with a realistic mixing model to compare reconstructed Mississippi River discharge against $\delta^{18}\text{O}_{\text{IVC-SW}}$ data from the Gulf of Mexico. We consider each of these steps, in turn.

Palaeotopographic and drainage network reconstruction. The topography over which we route flow through the Mississippi is a combination of modern surface elevation from the GEBCO_08 30-arcsecond global digital elevation model (<http://www.gebco.net>), perturbations to topography associated with GIA¹⁷, and ice thicknesses over glaciated areas^{3–6}. Using the evolving palaeotopography, we compute shorelines, flow paths (supraglacial, although these closely mirror subglacial drainage), and drainage basins using the Geographical Resources Analysis Support System (GRASS), an open-source Geographic Information Systems (GIS) software package^{31,32} and its drainage-solving algorithm, *r.watershed*³³. Performing flow-routing calculations at 30-arcsecond resolution is essential to resolve river valleys: computations on coarser grids can, for example, send the Upper Missouri River towards the wrong catchment.

Changes in drainage basin geometry are significantly more sensitive to the ice-sheet geometry than to GIA because the lower density of ice (917 kg m^{-3}) relative to mantle material ($3,300 \text{ kg m}^{-3}$) limits the isostatic effect on surface topography to about 30% of ice-sheet thickness. Drainage patterns become sensitive to GIA when ice-sheet retreat outpaces isostatic adjustment. In spite of this, the difference between sharp drainage boundaries and the low-pass filtering effect of lithospheric strength on flexural isostasy means that drainage reconstructions coupled with $\delta^{18}\text{O}_{\text{IVC-SW}}$ inversions provide a set of constraints on ice-sheet evolution that is largely different from those imposed by sea level markers and uplift rates that record the GIA process^{20,34}.

Our flow-routing calculations do not include the effects of erosion and deposition in modifying the landscape for two major reasons. First, evidence from Lake Agassiz drainage rearrangements show that drainages can change direction after spillways have been cut²⁷, meaning that ice sheets and GIA act as a primary driver of drainage change. Second, we want to compare our drainage histories with observations, which requires that we avoid preconditioning our model to match these observations.

Our glacial isostatic adjustment (GIA) calculations are based on a gravitationally self-consistent sea-level theory that incorporates viscoelastic deformation of the Earth, time-varying shorelines, changes in the extent of grounded marine-based ice, and perturbations in sea level due to variations in the Earth's rotation^{17,35}. The calculations are performed using a pseudo-spectral algorithm truncated at degree and order 256. This resolution is sufficient, given the adopted models of the lithosphere (see below), to ensure convergence.

The GIA simulations are based on four different published ice histories. Three of these ice histories, ICE-3G³, ICE-5G⁵ and the ANU model⁴ are global models, and in each case the GIA calculations require that the ice model be paired with the viscoelasticity model used in its derivation: VM1 for ICE-3G; VM2 for ICE-5G; and a model characterized by an elastic lithosphere of thickness 70 km, and upper and lower mantle viscosities of $3 \times 10^{20} \text{ Pa s}$ and $5 \times 10^{21} \text{ Pa s}$, respectively, for the ANU model. The fourth ice model (G12) is constructed by combining the Cordilleran–Laurentide–Greenland ice-sheet complex derived using a dynamical ice model⁶ with additional ice volume taken from the ICE-5G sea-level history, and coupling it with the VM2 solid Earth model. The ICE-3G, ANU and G12 models are initiated at the LGM (or just before this time), whereas ICE-5G begins at the end of the last interglacial. Finally, in computing palaeotopography (but not in the GIA calculations), the ice models are interpolated onto smooth grids with 30-arcsecond resolution to match the resolution of the GEBCO_08 global digital elevation model (<http://www.gebco.net>).

River discharge. River discharge is a combination of two components. The first is seasonal water input from precipitation minus evapotranspiration. The second is ice-sheet melt.

Time-evolving precipitation and evapotranspiration fields were extracted from the TRACE-21K LGM-to-present continuous run of the Community Climate System Model version 3 (CCSM3)^{14,15}, smoothed and interpolated to 30-arcsecond resolution. These fields were corrected for model biases³⁶ by forcing the precipitation and evapotranspiration fields at 0 kyr BP to conform to modern precipitation reanalysis data¹³ and satellite-derived evapotranspiration products^{16,37}. As a result, CCSM3 provided the difference between past and modern climate. Meteoric water discharge down the Mississippi River was calculated by summing these inputs across its computed drainage basin at each time step. The resulting time series was scaled to match the modern Mississippi discharge at 0 kyr BP. Without this final correction, the computed modern Mississippi discharge would be about 20% higher than the measured mean value³⁸.

Meltwater discharge was obtained by computing the mass balance of the smoothed and interpolated ice models within the evolving Mississippi drainage basin. Where more positive $\delta^{18}\text{O}_{\text{IVC-SW}}$ values indicate water storage in the ice sheets, we subtracted this storage term from the volume of meteoric water inputs.

We neglect groundwater contributions because the maximum modelled deglacial groundwater net exfiltration discharge across the whole North American continent is only around 2.5% of the modern Mississippi River discharge^{38,39}. Likewise, we do not consider proglacial lakes because the ice-sheet reconstructions^{3–6} do not include them in their mass balance and GIA calculations, meaning that they are implicitly included in the ice volume.

Mixing model and $\delta^{18}\text{O}$ data. We compiled high-resolution $\delta^{18}\text{O}$ data sets from measurements of white *Globigerinoides ruber*, a planktic foraminifer, in Gulf of Mexico cores from the Orca^{7,8,10,11,40} and Pigmy⁹ basins to test our ice-reconstruction-based calculations. We selected only those data with Mg/Ca as a proxy for sea surface temperature (SST), and corrected them for both SST^{8–10} and global ice volume^{7,41} to isolate the effects of river inputs¹¹. We termed these corrected data $\delta^{18}\text{O}_{\text{IVC-SW}}$ with 'ivc' indicating that they are corrected for global ice volume, and 'sw' indicating that they provide a sea-water value that excludes temperature effects. Many of these data were sampled every 5 mm (near-decadal intervals), with radiocarbon ages gathered every 10 cm (intervals of a few tens to hundreds of years)^{8–11}. Neither the fidelity nor the Mg/Ca SST corrections that characterize these data sets were available for the Louisiana Slope data used previously to estimate meltwater discharges into the Gulf of Mexico¹⁹.

All radiocarbon ages and associated errors were recalibrated with IntCal09⁴², using published marine reservoir corrections and analytical errors^{7–11,40}, to create a homogeneous set of calendar ages. These were incorporated into the statistical model clam⁴³ to generate new age models and error distributions for each data set. In generating these age models, we considered the homogeneous interval in core MD02-2550⁴⁴, which contained two nearly identical radiocarbon ages 14 cm apart ($12,785 \pm 45 \text{ }^{14}\text{C}$ years and $12,800 \pm 40 \text{ }^{14}\text{C}$ years BP before the reservoir correction, placing both around 14,400–14,500 calendar years BP)¹⁰, to be an event bed.

To compare these data against synthetic meltwater discharge histories derived from the ice-sheet reconstructions^{3–6}, we adopted a mixing model¹² that describes $\delta^{18}\text{O}_{\text{IVC-SW}}$ in the Gulf of Mexico as a function of inputs from ice-sheet melt (subscript 'i'), precipitation and evapotranspiration that is not incorporated into the ice sheet (subscript 'p'), and ocean water from the Gulf of Mexico loop current (subscript 'lc'). Discharges (in units of cubic metres per second) are denoted by Q .

$$\delta^{18}\text{O}_{\text{IVC-SW}} = \frac{(\delta^{18}\text{O})_i Q_i + (\delta^{18}\text{O})_p Q_p + (\delta^{18}\text{O})_{lc} Q_{lc}}{Q_i + Q_p + Q_{lc}} \quad (1)$$

$\delta^{18}\text{O}_i$ ranges from -25‰ , found in groundwater in the Lake Agassiz basin²⁹, to -38‰ , from basal Pleistocene ice in the Barnes Ice Cap³⁰. This range defines the error envelopes shown in Fig. 2a. $\delta^{18}\text{O}_p$ is -7.7‰ , from modern measurements of Mississippi springtime runoff that are assumed to approximate glacial conditions¹⁹. $\delta^{18}\text{O}_{lc}$ is $+1.1\text{‰}$; this is based on modern measurements⁴⁵ and is consistent with the ice-volume-corrected value used in prior work¹⁹.

We first used equation (1) to compute the input discharge from the loop current (Q_{lc}), assuming, as in previous studies^{12,19,46,47}, that it is constant with time. This assumption is upheld over the last 10,000 years by the approximately constant $\delta^{18}\text{O}_{\text{IVC-SW}}$ values from Gulf of Mexico core sections that post-date continental glaciation in the Mississippi drainage basin. Using $\delta^{18}\text{O}_{\text{IVC-SW}}$ data from the past 1,400 years⁹, the modern mean Mississippi discharge of $16,790 \text{ m}^3 \text{ s}^{-1}$ (ref. 38), and no meltwater inputs, we solve for $Q_{lc} = 596,000 \text{ m}^3 \text{ s}^{-1}$.

We next solved equation (1) to compute synthetic oxygen isotope curves for each ice model we analysed (Fig. 2a), and conversely to convert the $\delta^{18}\text{O}_{\text{IVC-SW}}$ data in the Gulf of Mexico into a data-driven meltwater discharge history (Figs 1 and 2b). The $\delta^{18}\text{O}_{\text{IVC-SW}}$ -based meltwater discharge history was calculated using an ice-sheet oxygen isotope ratio of $\delta^{18}\text{O}_i = -31.5\text{‰}$, which is the mean of the measured values^{29,30}. Seasonal precipitation minus evapotranspiration time series for the $\delta^{18}\text{O}_{\text{IVC-SW}}$ -based discharge history were calculated by summing the corrected^{13,16,36–38} general-circulation-model outputs^{14,15} across the time-evolving drainage basins calculated from each of the four ice models^{3–6} (Fig. 3).

Gulf of Mexico loop current. A critical component to any water mixing model in the Gulf of Mexico is the input discharge from eddy cut-offs of the Gulf of Mexico loop current. The basis for all prior approximations of loop-current discharge used in water mixing models is a geometric approximation of the size of loop-current eddies by ref. 46. This approximation was parsed into surface- and deep-water components¹⁹ for more recent work^{12,19}. The surface-water loop-current discharge estimate, which is of relevance to us in our work with the planktic foraminifer *G. ruber*, is $159,000 \text{ m}^3 \text{ s}^{-1}$.

Previous studies focused only on MWP-1A, but here we compiled data that continue to the modern day⁹. This allowed us to back-calculate the loop-current

discharge by combining known modern Mississippi River discharge (ref. 38) with postglacial measurements of $\delta^{18}\text{O}_{\text{IVC-SW}}$ (ref. 9) in a water mixing model (equation (1) and Fig. 2). The result of this calculation is $Q_{\text{IC}} = 596,000 \text{ m}^3 \text{ s}^{-1}$, nearly four times the value used previously^{12,19}. This difference highlights the importance of carrying the analyses through to the modern day, and shows that previous meltwater discharge estimates were inconsistent with modern data. Our higher loop-current discharge means that Mississippi River discharge will have a smaller effect on the mixed $\delta^{18}\text{O}_{\text{IVC-SW}}$ and therefore that negative $\delta^{18}\text{O}_{\text{IVC-SW}}$ excursions represent a much larger increase in meltwater discharge than previously^{12,19} thought.

Precipitation inputs. Seasonal meteoric water (precipitation minus evapotranspiration) inputs are also important to the results of the mixing model. These are strongly controlled by the time-evolving drainage basin area of the Mississippi River, and are modified by changing precipitation and evapotranspiration rates. We calculate around $54,000 \text{ m}^3 \text{ s}^{-1}$ in precipitation-derived runoff during MWP-1A, averaged across the drainage reconstructions based on all four ice models. This is much greater than the previous estimates of $30,000 \text{ m}^3 \text{ s}^{-1}$ (ref. 12) and $18,000 \text{ m}^3 \text{ s}^{-1}$ (ref. 19) used in mixing models.

Hyperpycnal discharge. Rivers typically discharge their water as hypopycnal (surface-water) plumes owing to the lower density of freshwater compared to sea water^{48,49}. In cases in which suspended sediment concentrations are high enough that the combined water–sediment density becomes greater than the density of sea water, water from the river instead dives towards the bed, becoming a hyperpycnal (bottom-water) plume.

Our analysis excludes any hyperpycnal component of meltwater discharge, which has been thought to be important during MWP-1A^{12,19,21}. Although benthic data from the Louisiana Slope have been used in prior studies to argue both for¹⁹ and against¹² a large Mississippi-routed contribution to MWP-1A, these data have sparse age control through MWP-1A that precludes acknowledgement of time-variable sedimentation rates. This results in a significant overestimation of hyperpycnal meltwater discharge to the Gulf of Mexico owing to the linear interpolation of time through an event bed.

A new high-resolution age model¹⁰ from Gulf of Mexico core MD02-2550 shows that during MWP-1A¹⁸, a massive 19-cm bed⁴⁴ was deposited essentially instantaneously. This matches the timing of the incision of the lower Mississippi River⁵⁰. Because sediment transport⁵¹ and geomorphic change⁵² are thresholded and strongly nonlinear, we infer that a brief, massive flooding event around the time of MWP-1A caused rapid onshore erosion and offshore deposition. This would be consistent with a brief period of hyperpycnal flow, because in spite of their rarity, these bottom-hugging plumes produce a disproportionately large amount of the stratigraphic record⁵³.

The literature on hyperpycnal flows shows that they are typically rare events associated with large floods in small catchments^{48,53–55}. The modern Mississippi, like many large rivers, is incapable of generating hyperpycnal flows⁵⁵ for two major reasons. First, most large modern rivers meet the ocean at broad, flat coastal plains. These low riverbed slopes cause sediment to fall from suspension, reducing its concentration to below the threshold for hyperpycnal flow initiation^{54,55}. Second, flooding is asynchronous throughout the drainage basin, and sharp flood waves from far upstream often disperse by the time they reach the ocean. Sediments derived from these floods are deposited on their way to the shore or diluted by water from the remainder of the basin, which is not experiencing flood conditions⁵⁵.

The deglacial Mississippi River flowed through canyons on the continental shelf, preventing nearshore sediment trapping, and was fed by occasional intense outburst floods²⁷, which could rapidly supply water and sediment to the coast. Therefore, the deglacial Mississippi may have been capable of producing hyperpycnal flows during extreme flood events.

Previous studies^{21,56,57} have stated that the Mississippi may have discharged a sustained hyperpycnal flow through MWP-1A that contributed significantly to global sea level. They cite a laboratory study in which sediment-laden portions of hypopycnal plumes mixed downward and became hyperpycnal at mean suspended sediment concentrations as low as 1 kg m^{-3} , much lower than the 40 kg m^{-3} typically required for wholesale hyperpycnal plume initiation⁵⁸. Although the laboratory study showed that downward mixing of the hypopycnal plume occurred, the bulk of the plume remained at the surface⁵⁸. Furthermore, Parsons *et al.*⁵⁸ note significant difficulties in scaling their laboratory experiment to the field, and that their observed downward mixing has not been observed in nature. These factors are in direct conflict with the interpretation of ref. 58 as supporting a wholesale sequestration of the Mississippi freshwater plume into deep water⁵⁷.

We therefore propose the following chain of events. An extreme deglacial flood first eroded the lower Mississippi valley⁵⁰. This sediment-laden water then produced a hyperpycnal plume that rapidly deposited the observed massive bed in core MD02-2550⁴⁴ as well as a thick layer of sediments in LOUIS core 2023¹⁹. Benthic foraminifera (*Uvigerina peregrina*) in the latter recorded a negative oxygen isotope excursion¹⁹ due to the hyperpycnal freshwater input that included the

highly negative $\delta^{18}\text{O}$ signature of ice-sheet melt. In spite of the geomorphic and sedimentary impact of such an event, it would not last long enough to contribute significantly to global sea level. As a result, we discount a persistent hyperpycnal flow as a significant contributor to total Mississippi meltwater discharge during MWP-1A^{12,19,21,56,57}, and attribute previous interpretations of it^{12,19} to an assumed constant deposition rate between too-widely-spaced radiocarbon ages that caused a thick, rapidly deposited event bed to appear to represent the long-term state of the system.

Laurentide–Cordilleran saddle collapse and G12 model offset. In the study that produced the G12 ice model⁶, the authors observed a saddle collapse between the Laurentide and Cordilleran ice sheets at 11.6 kyr BP. They offset the model by 2.5 kyr to move the peak of meltwater discharge associated with the saddle collapse to 14.1 kyr BP, which they take to be part of MWP-1A. Our work with their model uses the original, unshifted model ages.

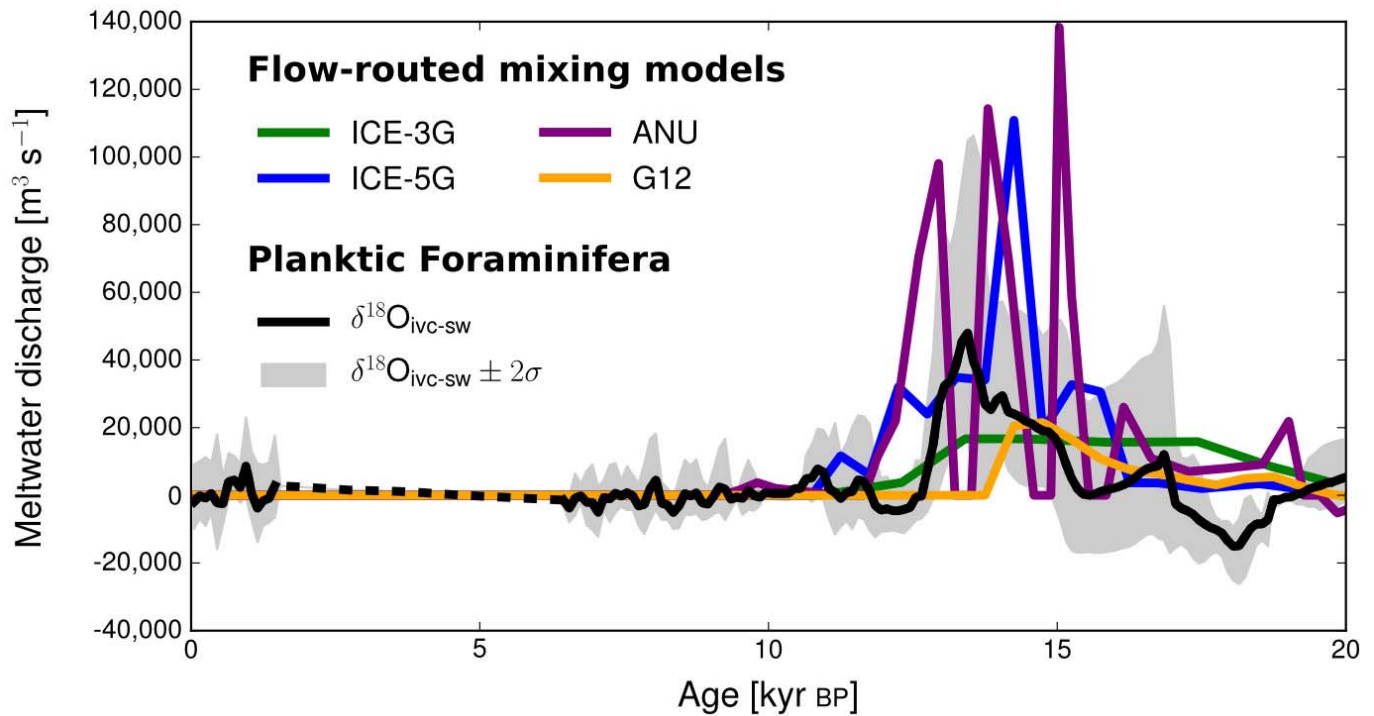
We used G12 as a representative ice-physics-based model. We sought to find whether such a model, without *a posteriori* modification, could match the oxygen isotope records in the Gulf of Mexico. This is essential to our broader goal of self-consistently integrating multiple data sets and models to generate improved ice-sheet histories. A model offset indicates that the climatic drivers, parametrized ice dynamics or basal conditions do not adequately allow the ice sheet to evolve and match data. By keeping model time and geologic time equivalent, future ice models can combine the constraints of ice dynamics and climate forcings with isotopic records, GIA, and other geological and geophysical constraints to build better reconstructions of the time-evolving Pleistocene ice-sheet complexes.

Making the G12 results 2.5 kyr older, as was previously done⁶, would cause its predicted $\delta^{18}\text{O}_{\text{IVC-SW}}$ history (Fig. 2a) to misfit the data strongly. Instead, we would have to shift G12 younger by 1.1 kyr to match the timing of its drainage shift away from the Mississippi with the observed flow re-routing at the onset of the Younger Dryas (about 12.9 kyr BP)^{11,24,27}. Because of this conflict—the saddle collapse at MWP-1A requiring the model to be shifted earlier in time⁶, and the drainage re-routing requiring that it be shifted later—it seems that the G12 model cannot conclusively implicate the saddle collapse between the Laurentide and Cordilleran ice sheets as a significant contributor to MWP-1A. Nevertheless, it is encouraging that this ice-physics-based model can match the geologic constraints on meltwater discharge in our study: given that it was not calibrated to these, they provide an external check on its accuracy.

When the Laurentide–Cordilleran saddle did collapse in G12 at 11.6 kyr BP, we see no meltwater input to the Mississippi (Fig. 2) in our drainage calculations. Our drainage calculations show that this saddle collapse instead routed flow through the Mackenzie and Saint Lawrence rivers. Likewise, we see that the Mackenzie River routes the runoff from the ice saddle collapse in our analysis of ICE-5G⁵. Whatever its timing, if this saddle collapse were as catastrophic as the G12 model indicates⁶, we should expect to see a major signature of it in the Mackenzie River delta and/or the Gulf of Saint Lawrence, as well as in the river systems that connected the ice saddle to the shore.

- Neteler, M., Bowman, M. H., Landa, M. & Metz, M. GRASS GIS: a multi-purpose open source GIS. *Environ. Modelling Software* **31**, 124–130 (2012).
- GRASS Development Team. *Geographic Resources Analysis Support System (GRASS GIS) Software* <http://grass.osgeo.org/> (Open Source Geospatial Foundation, 2012).
- Metz, M., Mitasova, H. & Harmon, R. S. Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrol. Earth Syst. Sci.* **15**, 667–678 (2011).
- Bassett, S. E., Milne, G. A., Mitrovica, J. X. & Clark, P. U. Ice sheet and solid Earth influences on far-field sea-level histories. *Quat. Sci. Rev.* **309** (2005).
- Mitrovica, J. X. & Milne, G. A. On post-glacial sea level: I. General theory. *Geophys. J. Int.* **154**, 253–267 (2003).
- Yeager, S. G., Shields, C. A., Large, W. G. & Hack, J. J. The low-resolution CCSM3. *J. Clim.* **19**, 2545–2566 (2006).
- Mu, Q., Heinsch, F. A., Zhao, M. & Running, S. W. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.* **111**, 519–536 (2007).
- Kammerer, J. C. *Largest Rivers in the United States* Open-File Report <http://pubs.usgs.gov/of/1987/ofr87-242/> (United States Geological Survey, 1990).
- Lemieux, J.-M., Sudicky, E. A., Peltier, W. R. & Tarasov, L. Dynamics of groundwater recharge and seepage over the Canadian landscape during the Wisconsinian glaciation. *J. Geophys. Res.* **113**, 1–18 (2008).
- Leventer, A., Williams, D. F. & Kennett, J. P. Dynamics of the Laurentide ice sheet during the last deglaciation: evidence from the Gulf of Mexico. *Earth Planet. Sci. Lett.* **59**, 11–17 (1982).
- Waelbroeck, C., Labeyrie, L., Michel, E., Duplessy, J. C. & McManus, J. F. Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **21**, 295–305 (2002).
- Reimer, P. *et al.* IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51**, 1111–1150 (2011).
- Blauw, M. Methods and code for ‘classical’ age-modelling of radiocarbon sequences. *Quat. Geochron.* **5**, 512–518 (2010).

44. Meckler, A. *et al.* Glacial to Holocene terrigenous organic matter input to sediments from Orca Basin, Gulf of Mexico—a combined optical and biomarker approach. *Earth Planet. Sci. Lett.* **272**, 251–263 (2008).
45. Wagner, A. J. & Slowey, N. C. Oxygen isotopes in seawater from the Texas-Louisiana Shelf. *Bull. Mar. Sci.* **87**, 1–12 (2011).
46. Emiliani, C., Rooth, C. & Stipp, J. J. The late Wisconsin flood into the Gulf of Mexico. *Earth Planet. Sci. Lett.* **41**, 159–162 (1978).
47. Aharon, P. Meltwater flooding events in the Gulf of Mexico revisited: implications for rapid climate changes during the last deglaciation. *Paleoceanography* **18**, 1079 (2003).
48. Kettner, A. J. & Syvitski, J. P. M. HydroTrend v. 3.0: a climate-driven hydrological transport model that simulates discharge and sediment load leaving a river system. *Comput. Geosci.* **34**, 1170–1183 (2008).
49. Lamb, M. P. & Mohrig, D. Do hyperpycnal-flow deposits record river-flood dynamics? *Geology* **37**, 1067–1070 (2009).
50. Marchitto, T. M. & Wei, K. Y. History of Laurentide meltwater flow to the Gulf of Mexico during the last deglaciation, as revealed by reworked calcareous nannofossils. *Geology* **23**, 779 (1995).
51. Wilcock, P. R. & Crowe, J. C. Surface-based transport model for mixed-size sediment. *J. Hydraulic Eng.* **129**, 120–128 (2003).
52. Wolman, M. G. & Miller, J. P. Magnitude and frequency of forces in geomorphic processes. *J. Geol.* **68**, 54–74 (1960).
53. Syvitski, J. & Kettner, A. J. On the flux of water and sediment into the Northern Adriatic Sea. *Continental Shelf Res.* **27**, 296–308 (2007).
54. Mulder, T. & Syvitski, J. P. M. Turbidity currents generated at river mouths during exceptional discharges to the world oceans. *J. Geol.* **103**, 285–299 (1995).
55. Mulder, T., Syvitski, J. P., Migeon, S., Faugères, J.-C. & Savoye, B. Marine hyperpycnal flows: initiation, behaviour and related deposits. *Mar. Petrol. Geol.* **20**, 861–882 (2003).
56. Tarasov, L. & Peltier, W. R. Arctic freshwater forcing of the Younger Dryas cold reversal. *Nature* **435**, 662–665 (2005).
57. Peltier, W. Rapid climate change and Arctic Ocean freshening. *Geology* **35**, 1147–1148 (2007).
58. Parsons, J. D., Bush, J. W. M. & Syvitski, J. P. M. Hyperpycnal plume formation from riverine outflows with small sediment concentrations. *Sedimentology* **48**, 465–478 (2001).



Extended Data Figure 1 | Meltwater discharge histories computed from each of the flow-routed ice models. For comparison, the $\delta^{18}\text{O}_{\text{ivc-sw}}$ data have been converted to meltwater discharge using the mixing model (equation (1)). The negative discharge shown in the data indicates that the LIS was growing

from precipitation inputs during the Port Bruce Readvance²⁴, significantly reducing net Mississippi discharge (precipitation minus evapotranspiration, minus ice sheet growth) during that time. The modern mean Mississippi discharge, for reference, is $16,790 \text{ m}^3 \text{ s}^{-1}$ (ref. 38).

Extended Data Table 1 | Ice-sheet and Mississippi drainage reconstructions

Ice history	$V_i(\text{NA})$	$\frac{A_{\text{db,LGM}}}{A_{\text{db,modern}}}$	0 kyr BP $\sum_{t=\text{LGM}} Q_i(t)$	14.31 kyr BP $\sum_{t=14.65 \text{ kyr BP}}^{14.31 \text{ kyr BP}} Q_i(t)$	Start time (approx. LGM)
	(m SLE)	(NA)	(m SLE)	(m SLE)	(yr BP)
ICE-3G ³	56.55	1.99	8.98	0.50	20,240
ICE-5G ⁵	87.98	1.68	15.66	3.30	20,000
ANU ⁴	90.71	1.96	19.35	1.73	20,000
G12 ⁶	94.23	1.94	4.52	0.61	20,000
$\delta^{18}\text{O}^{7-11}$	NA	NA	5.42 ± 2.09	0.66 ± 0.07	19,950

We calculated meltwater discharges from four ice models, as well as by inverting $\delta^{18}\text{O}$ data using a mixing model (equation (1)). $V_i(\text{NA})$ is the initial ice volume for each ice model over our study region, which covers all of the North American mainland and approximately the western half of Greenland. $(A_{\text{db,LGM}})/(A_{\text{db,modern}})$ is the LGM drainage area of the Mississippi River ($A_{\text{db,LGM}}$), normalized to the modern drainage area ($A_{\text{db,modern}}$). $\sum_{t=\text{LGM}}^{0 \text{ kyr BP}} Q_i(t)$ is the total meltwater contribution to global sea level rise routed down the Mississippi River. $\sum_{t=14.65 \text{ kyr BP}}^{14.31 \text{ kyr BP}} Q_i(t)$ is the meltwater volume routed down the Mississippi during the 14.65–14.31-kyr-BP¹⁸ MWP-1A window. Meltwater volumes can be converted to SI units by multiplying the SLE value by the approximate ocean surface area $3.61 \times 10^{14} \text{ m}^2$. Start time is the starting time step used in our analysis, which is shortly after the canonical 21-kyr BP LGM because of the differencing between ice-model time steps that is necessary to estimate meltwater discharges. However, because these models include little to no ice retreat between 21 kyr BP and 20 kyr BP, the error introduced by starting at 20 kyr BP is negligible. NA, not applicable.

Decoupling of soil nutrient cycles as a function of aridity in global drylands

Manuel Delgado-Baquerizo^{1,2}, Fernando T. Maestre², Antonio Gallardo¹, Matthew A. Bowker³, Matthew D. Wallenstein⁴, Jose Luis Quero^{2,5}, Victoria Ochoa², Beatriz Gozalo², Miguel García-Gómez², Santiago Soliveres², Pablo García-Palacios^{4,6}, Miguel Berdugo², Enrique Valencia², Cristina Escolar², Tulio Arredondo⁷, Claudia Barraza-Zepeda⁸, Donaldo Bran⁹, José Antonio Carreira¹⁰, Mohamed Chaieb¹¹, Abel A. Conceição¹², Mchich Derak¹³, David J. Eldridge¹⁴, Adrián Escudero², Carlos I. Espinosa¹⁵, Juan Gaitán⁹, M. Gabriel Gatica¹⁶, Susana Gómez-González¹⁷, Elizabeth Guzman¹⁵, Julio R. Gutiérrez⁸, Adriana Fiorentino¹⁸, Estela Hepper¹⁹, Rosa M. Hernández²⁰, Elisabeth Huber-Sannwald⁷, Mohammad Jankju²¹, Jushan Liu²², Rebecca L. Mau²³, Maria Miriri²⁴, Jorge Monerris²⁵, Kamal Naseri²¹, Zouhaier Noumi¹¹, Vicente Polo², Anibal Prina¹⁹, Eduardo Pucheta¹⁶, Elizabeth Ramírez²⁰, David A. Ramírez-Collantes²⁶, Roberto Romão¹², Matthew Tighe²⁷, Duilio Torres²⁸, Cristian Torres-Díaz¹⁷, Eugene D. Ungar²⁹, James Val³⁰, Wanyoike Wamiti³¹, Deli Wang²² & Eli Zaady³²

The biogeochemical cycles of carbon (C), nitrogen (N) and phosphorus (P) are interlinked by primary production, respiration and decomposition in terrestrial ecosystems¹. It has been suggested that the C, N and P cycles could become uncoupled under rapid climate change because of the different degrees of control exerted on the supply of these elements by biological and geochemical processes^{1–5}. Climatic controls on biogeochemical cycles are particularly relevant in arid, semi-arid and dry sub-humid ecosystems (drylands) because their biological activity is mainly driven by water availability^{6–8}. The increase in aridity predicted for the twenty-first century in many drylands worldwide^{9–11} may therefore threaten the balance between these cycles, differentially affecting the availability of essential nutrients^{12–14}. Here we evaluate how aridity affects the balance between C, N and P in soils collected from 224 dryland sites from all continents except Antarctica. We find a negative effect of aridity on the concentration of soil organic C and total N, but a positive effect on the concentration of inorganic P. Aridity is negatively related to plant cover, which may favour the dominance of physical processes such as rock weathering, a major source of P to ecosystems, over biological processes that provide more C and N, such as litter decomposition^{12–14}. Our findings suggest that any predicted increase in aridity with climate change will probably reduce the concentrations of N and C in global drylands, but increase that of P. These changes would uncouple the C, N and P cycles in drylands and could negatively affect the provision of key services provided by these ecosystems.

Biogeochemical cycles are biologically coupled, on molecular to global scales, owing to the conserved elemental stoichiometry of plants and microorganisms that drive the cycling of C, N and P (ref. 1). The availability of C and N is primarily linked to biological processes such as photosynthesis, atmospheric N fixation and subsequent microbial mineralization^{12–14}. However, available P for plants and microorganisms^{2,13,14} is derived mainly from mechanical rock weathering and, to a lesser extent, from the decomposition of organic matter^{12–14}. The importance of biological control of nutrient cycling relative to geochemical control has been shown to change with ecosystem development^{2,13,14}. For example, during the earliest stages of ecosystem succession, a relative prevalence of geochemical control on nutrient cycling means that P is made available by mechanical rock weathering, but that N and C are scarce, leading to a disparity in the C, N and P cycles relative to plant nutrient requirements^{13–16}. Climatic controls on ecosystem development and biogeochemical cycles are particularly relevant in drylands because their biological activity is mainly driven by water availability^{6–8}. Drylands cover about 41% of Earth's land surface and support more than 38% of the global human population¹⁷, constituting the largest terrestrial biome on the planet¹⁸. The increase in aridity predicted for the late-twenty-first century in many regions worldwide will increase the total area of drylands globally^{9–11}. These changes are predicted to exacerbate processes leading to land degradation and desertification, which already threaten the livelihood of more than 250 million people living in drylands^{17,18}. For example, a worldwide decrease in soil moisture by 5–15% has been predicted for the 2080–2099 period¹¹. Of particular concern is that the

¹Departamento de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, Carretera de Utrera, kilómetro 1, 41013 Sevilla, Spain. ²Área de Biodiversidad y Conservación, Departamento de Biología y Geología, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Calle Tulipán Sin Número, 28933 Móstoles, Spain. ³School of Forestry, Northern Arizona University, Flagstaff, Arizona 86011, USA. ⁴Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado 80523, USA. ⁵Departamento de Ingeniería Forestal, Campus de Rabanales Universidad de Córdoba, Carretera de Madrid, kilómetro 396, 14071 Córdoba, Spain. ⁶Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA. ⁷División de Ciencias Ambientales, Instituto Potosino de Investigación Científica y Tecnológica, San Luis Potosí, San Luis Potosí, 78210, Mexico. ⁸Departamento de Biología, Universidad de La Serena, La Serena 599, 1700000, Chile. ⁹Instituto Nacional de Tecnología Agropecuaria, Estación Experimental San Carlos de Bariloche 277, Bariloche, Río Negro, 8400, Argentina. ¹⁰Universidad de Jaén, Departamento de Biología Animal, Biología Vegetal y Ecología, 23071 Jaén, Spain. ¹¹Université de Sfax, Faculté des Sciences, Unité de Recherche Plant Diversity and Ecosystems in Arid Environments, Route de Sokra, kilomètre 3.5, Boîte Postale 802, 3018 Sfax, Tunisia. ¹²Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Avenida Transnordestina Sin Número, Bairro Novo Horizonte, Feira de Santana, 44036-900, Brasil. ¹³Direction Régionale des Eaux et Forêts et de la Lutte Contre la Désertification du Rif, Avenue Mohamed 5, Boîte Postale 722, 93000 Tétouan, Morocco. ¹⁴School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia. ¹⁵Instituto de Ecología, Universidad Técnica Particular de Loja, San Cayetano Alto, Marcelino Champagnat, Loja, 11-01-608, Ecuador. ¹⁶Departamento de Biología, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan, Rivadavia, San Juan, J5402DCS, Argentina. ¹⁷Laboratorio de Genómica y Biodiversidad, Departamento de Ciencias Básicas, Universidad del Bío-Bío 447, Chillán, 3780000, Chile. ¹⁸Instituto de Edafología, Facultad de Agronomía, Universidad Central de Venezuela, Ciudad Universitaria, Caracas, 1051, Venezuela. ¹⁹Facultad de Agronomía, Universidad Nacional de La Pampa, Casilla de Correo 300, 6300 Santa Rosa, La Pampa, Argentina. ²⁰Laboratorio de Biogeoquímica, Centro de Agroecología Tropical, Universidad Experimental Simón Rodríguez, Caracas, 47925, Venezuela. ²¹Department of Range and Watershed Management, Faculty of Natural Resources and Environment, Ferdowsi University of Mashhad, Azadi Square, Mashhad 91775–1363, Iran. ²²Institute of Grassland Science, Northeast Normal University and Key Laboratory of Vegetation Ecology, Ministry of Education, Changchun, Jilin Province 130024, China. ²³Department of Biological Sciences, Northern Arizona University, PO Box 5640, Flagstaff, Arizona 86011–5640, USA. ²⁴Department of Evolution, Ecology and Organismal Biology, Ohio State University, 318 West 12th Avenue, Columbus, Ohio 43210, USA. ²⁵Université du Québec à Montréal Pavillon des Sciences Biologiques, Département des Sciences Biologiques, 141 Président-Kennedy, Montréal, Québec H2X 3Y5, Canada. ²⁶Production Systems and the Environment Sub-Program, International Potato Center. Apartado 1558, Lima 12, Peru. ²⁷Department of Agronomy and Soil Science, School of Environmental and Rural Science, University of New England, Armidale, New South Wales 2351, Australia. ²⁸Departamento de Química y Suelos, Decanato de Agronomía, Universidad Centroccidental “Lisandro Alvarado”, Barquisimeto 3001, Venezuela. ²⁹Department of Agronomy and Natural Resources, Institute of Plant Sciences, Agricultural Research Organization, The Volcani Center, Bet Dagan 50250, Israel. ³⁰Office of Environment and Heritage, PO Box 363, Buronga, New South Wales 2739, Australia. ³¹Zoology Department, National Museums of Kenya, Ngara Road, Nairobi, 78420-00500, Kenya. ³²Department of Natural Resources and Agronomy, Agriculture Research Organization, Ministry of Agriculture, Gilat Research Center, Mobile Post Negev 85280, Israel.

biogeochemical cycles of C, N and P could become uncoupled under rapid climate change because of the different degrees of control exerted on these elements by biological and geochemical processes^{1–5}. As the global human population continues to grow, we will rely increasingly on marginal lands—particularly drylands—for the production of food, wood and biofuels, and to offset the emission of greenhouse gases^{3,4,17,18}. These ecosystem services can be greatly and negatively affected by the decoupling of the biogeochemical cycles of C, N and P in soils^{3,4,17,18}. Despite the importance of these cycles for ecosystem functioning and human welfare, it is largely unknown how predicted increases in aridity may influence them^{4,18}, and no global field studies have yet been conducted on this topic¹⁹.

We evaluated how aridity affects the balance between C, N and P in soils collected from 224 dryland sites from all continents except Antarctica. Because aridity is a fundamental driver of biological and geochemical processes in drylands^{8,12,17}, we predicted that increasing aridity would reduce biological activity^{4,5} and, therefore, the availability in nutrients under more strict biological control³ (C and N), but would favour the relative dominance of nutrients linked to geochemical processes^{1–4,13,20} (P), causing a stoichiometric imbalance in the nutrient cycles associated with C and N (ref. 3). We selected organic C, total N and total P as

surrogates for C, N and P availability because they were highly related to other available C, N and P forms for plants and microorganisms such as dissolved carbohydrates, amino acids, inorganic N, Olsen inorganic P and HCl-P (fraction of P linked to calcium carbonate minerals) (Methods). Negative quadratic relationships were observed between aridity and both organic C and total N concentrations (Fig. 1a, c). Although nonsignificant, a positive trend was observed between aridity and total P (Fig. 1e). This relationship was significant when inorganic P was considered instead of total P (Extended Data Fig. 1a). Likewise, a negative quadratic relationship was observed between aridity and the N:P and C:P concentration ratios (Fig. 1b, d and Extended Data Fig. 1b, c). The C:N ratio decreased linearly with increasing aridity (Fig. 1f). Similar results were found when evaluating relationships between aridity and more labile C (carbohydrates), N (sum of dissolved inorganic N and amino acids) and P (available P) fractions, as well as with their respective C, N and P ratios (Extended Data Fig. 2). Mechanical rock weathering is the main P input into terrestrial ecosystems, but N is either absent from or uncommon in primary minerals, and inputs therefore are largely derived from atmospheric N fixation, deposition or both^{13,14}. Although rates of biological weathering should decrease with increasing aridity, mechanical rock weathering may increase with aridity, releasing P-bearing

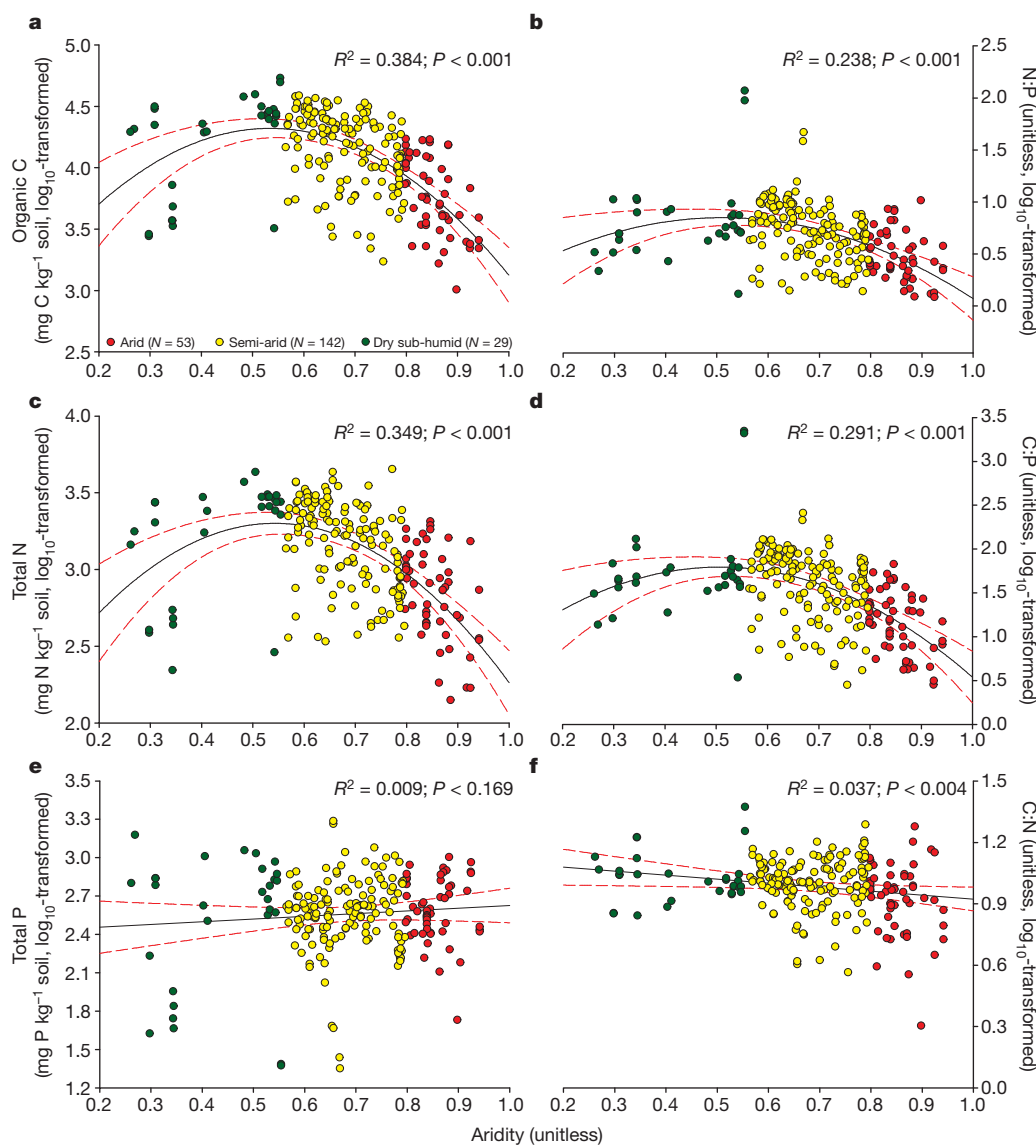


Figure 1 | Relationships between aridity and C, N and P at our study sites. a, Organic C; b, N:P; c, total N; d, C:P; e, total P; f, C:N. Aridity is defined as $1 - AI$, where AI, the ratio of precipitation to potential evapotranspiration,

is the aridity index. The solid and dashed lines represent the fitted quadratic regressions and their 95% confidence intervals, respectively. R^2 , proportion of variance explained.

minerals^{14,18,21}. Furthermore, the reduced plant activity and nutrient uptake typically observed in the most arid sites can also promote a higher availability of P at these sites^{2,7,12,14}. In addition, we found a positive relationship between aridity and the concentration of HCl-P (Extended Data Fig. 3). These results suggest that at least part of the increase in P derived from increasing aridity may be associated with the calcium carbonate minerals. High concentrations of calcium carbonates are likely to occur in the most arid soils because the low rainfall and high evaporation prevent weathering products from being washed out of the soil profile²⁰.

To clarify the effects and relative importance of aridity on the availability of C, N and P, we generated a structural equation model based on the known effects and relationships between key drivers of organic C, total N and total P (Extended Data Fig. 4). We included in this model phosphatase activity, which is the enzyme responsible for releasing inorganic P from organic sources and is considered a surrogate of biological P demand^{15,22,23}. Our model explained 43%, 26% and 53% of the variance in the organic matter component (first component from a principal-component analysis (PCA) conducted with organic C and total N), total P and phosphatase activity, respectively. Aridity had a direct negative effect on the organic matter component and phosphatase activity, but a positive effect on total P (Fig. 2). In addition, aridity was also the most important predictor of the organic matter component and phosphatase activity (Fig. 3). Similar results were found when we used inorganic P instead of total P in this model (Extended Data Figs 5 and 6) and when we included a more labile organic matter component (first component from a PCA with carbohydrates and available N) and available P (Extended Data Figs 7 and 8).

Our results imply a set of predictions. Forecasted increases in aridity in drylands globally^{9–11} are expected to lead to severe nutrient depletion¹⁴ in these environments, particularly in the most arid sites. For example, the observed decrease in N in drylands with increasing aridity (probably derived from the higher soil erosion and the lower plant cover), which

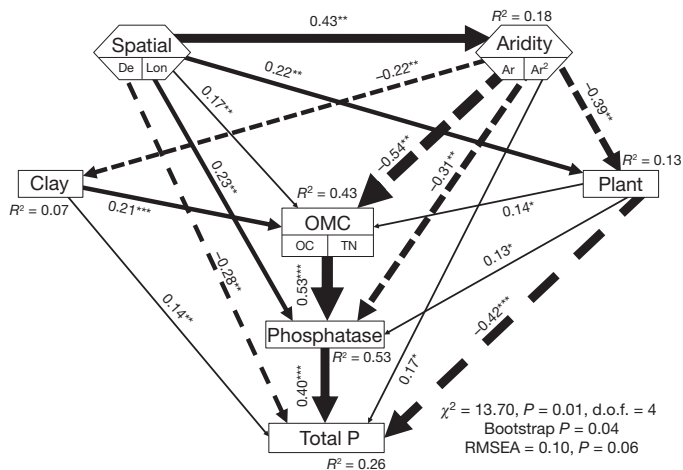


Figure 2 | Effects of aridity, clay percentage, plant cover and site position on the organic matter component, total P concentration and phosphatase activity. Spatial coordinates of the study sites are expressed in terms of distance from Equator (De) and longitude (Lon). Numbers adjacent to arrows are standardized path coefficients, analogous to relative regression weights, and indicative of the effect size of the relationship. Continuous and dashed arrows indicate positive and negative relationships, respectively. Arrow width is proportional to the strength of the relationship. The proportion of variance explained (R^2) appears alongside every response variable in the model. Goodness-of-fit statistics for each model are shown in the lower right corner (d.o.f., degrees of freedom; RMSEA, root mean squared error of approximation). There are some differences between the a-priori model and the final model structures, owing to removal of paths with coefficients close to zero (see the a-priori model in Extended Data Fig. 4). Hexagons are composite variables³⁰. Squares are observable variables. The organic matter component (OMC) is the first component from a PCA conducted with soil organic carbon (OC) and total nitrogen (TN). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

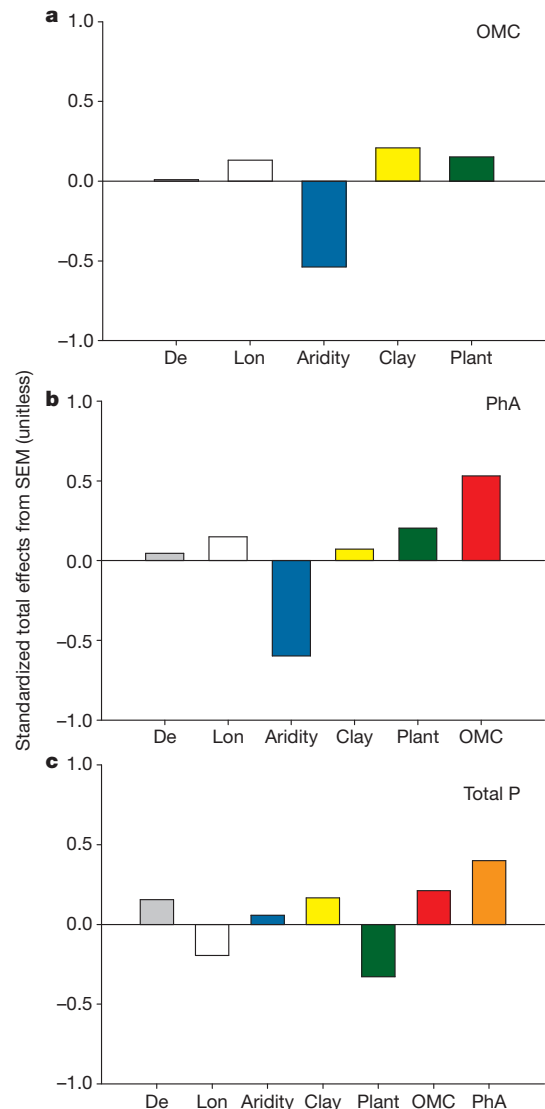


Figure 3 | Standardized total effects (direct plus indirect effects) derived from the structural equation modelling. These include the effects of aridity, percentage of clay, plant cover, distance from Equator (De) and longitude (Lon) on the organic matter component (OMC; first component from a PCA conducted with soil organic C and total N), total P and phosphatase activity (PhA). SEM, structural equation model.

are already poor in nutrients^{4,12}, could inhibit N mineralization in soils, potentially leading to a positive feedback on nutrient availability²⁴ (Extended Data Figs 9 and 10). The observed increase in the N:P ratio with decreasing aridity is similar to what would be expected during long-term ecosystem development^{13,14}. Although this progression has been described on the geological timescale (thousands to millions of years), and changes in aridity occur on the ecological timescale (tens to thousands of years), the processes may share the same biogeochemical signatures, such as N and P accumulation mediated by shifts in the relative importance of biological and geochemical processes^{13,14}. Thus, in the more arid sites, inorganic P accumulates by geochemical weathering because of the low net primary production and nutrient plant uptake characterizing these ecosystems, and is associated with calcium carbonate minerals. However, N is only slowly incorporated by N fixation^{13,14}, being also limited by low C (energy) availability^{1,3}. In the less arid sites (dry sub-humid ecosystems), where N and C are already available for plants and microorganisms, P becomes available through the activity of extracellular phosphatase enzymes¹⁴ (which require N investment), coupling P availability to biological processes²⁴.

Predicted increases in aridity, and, hence, decreases in water availability, are also expected to reduce plant cover in drylands⁵ (Fig. 2), favouring the dominance of physical processes²¹ (for example the abrasion of exposed rock surfaces by wind-blown sands) over biological processes (for example litter decomposition and N fixation), decreasing the concentration of N and C (for example because of the wind erosion) and increasing the concentration of P (for example because of bedrock rejuvenation) in soils, hence distorting soil C, N and P cycles^{1,21}. Carbon and N become uncoupled from P in response to increasing aridity (Fig. 1). Under arid and semi-arid conditions, the reduced availability of C and N may unbalance the concentrations of C, N and P, constraining plant and microbial activity and diversity^{1,3}. This may have an important negative effect on primary production and organic matter decomposition^{1,18,24}, even when P is available. In addition, the observed decrease in phosphatase activity and the increase in P observed in coarser, sandier soils with increasing aridity suggests that key abiotic and biotic processes, such as soil formation and organic matter decomposition, may be reduced with increasing aridity²¹. The decrease in the C:N ratio with increasing aridity observed here accords with experimental studies showing that drought periods decouple C and N cycles in drylands²⁵. Although organic C and total N are strongly correlated in the studied sites (Pearson's $r = 0.92$), our results suggest that future climatic conditions will promote N losses in drylands, particularly if increases in aridity reduce vegetation cover in these ecosystems (Fig. 2). The imbalance observed in the C, N and P cycles with increasing aridity may have other important consequences for drylands. For example, reductions in N availability as a consequence of increases in aridity will limit the capacity of plant primary productivity to buffer human-induced increases in atmospheric CO₂ concentrations, because the rate of photosynthesis is proportional to the amount and activity of the N-rich enzyme ribulose biphosphate carboxylase/oxygenase in leaves^{1,25}. This would contribute to a warmer world by the end of the twenty-first century, by limiting the capacity of plants and microorganisms to fix CO₂ derived from human activities^{1,3,26}. In addition, decreases in the supply of N relative to that of P may have a short-term effect by differentially constraining the growth rates of plant species based on their stoichiometry³. Such reductions may also have long-term evolutionary effects by selecting plants and microorganisms with different levels of N in their nucleotides, potentially altering ecosystem structural and functional traits³.

Our results indicate that the coupling between biogeochemical cycles in drylands will be particularly fragile in the face of rapid climate change, especially in the areas of transition between semi-arid and arid climates. Carbon, N and P availability seemed to be more resistant to changes in aridity in the transition from dry sub-humid to semi-arid climates than from semi-arid to arid, where we observed substantial and abrupt declines in organic C and total N, but an increase in inorganic P (Fig. 1a, c; Extended Data Fig. 1a). Similarly, we observed an abrupt decrease in the N:P and C:P ratios between semi-arid and arid sites, which was not observed in the C:N ratio (Fig. 1b, d, f). Evaluation of critical global transitions and tipping points is of major importance in assessing the effects of global change on ecosystems, and is an area of active research²⁷. The abrupt changes observed in the C:P and N:P ratios in the transitions from semi-arid to arid climates, together with the predicted increase in the proportion of global drylands considered to be arid⁹, may force these systems into a long process towards the recovery of ecosystem stoichiometry.

Our findings suggest that the predicted increase in aridity across drylands worldwide will reduce the concentration of biologically controlled C and N, but will increase that of P, which is primarily derived from rock weathering. These changes are likely to interrupt the C, N and P cycles in drylands in a nonlinear manner, and will have negative effects on biogeochemical reactions controlling key ecosystem functions (for example primary production, respiration and decomposition) and services (for example food production and carbon storage) from local to global scales¹.

METHODS SUMMARY

Field data were collected from 224 dryland sites located in 16 countries from all continents except Antarctica. At each site, to measure the total cover of perennial plants we established a 30 m × 30 m plot representative of the dominant vegetation²⁸. Five composite samples (0–7.5-cm depth) were randomly taken under the canopy of the dominant perennial plant species and in open areas devoid of perennial vegetation. The percentage of perennial plant cover; the percentage of clay; the concentrations of organic C, total N and total P; and the activity of phosphatase were determined as described in ref. 28. All these variables were then averaged to obtain site-level estimates by using the mean values observed in bare ground and vegetated areas, weighted by their respective cover at each site.

We first explored the relationship between aridity and organic C, total N, total P and the N:P, C:P and N:C ratios by using either linear or curvilinear regressions. We estimated the aridity¹⁹ (1 – AI) of each site using data from the WorldClim global database²⁹. We then used structural equation modelling³⁰ to examine the relative importance of aridity for organic C, total N, total P and phosphatase activity. Because organic C and total N were very closely related (Pearson's $r = 0.92$), we reduced these two variables to a single variable using PCA of the correlation matrix. We then introduced the first component of this PCA as a new variable into the model (organic matter component).

We evaluated the fit of our model using the model χ^2 -test and the root mean squared error of approximation; because the residuals of some data were not normally distributed, we confirmed fit using the Bollen–Stine bootstrap test (Fig. 2).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 March; accepted 17 September 2013.

1. Finzi, A. C. *et al.* Coupled biochemical cycles: responses and feedbacks of coupled biogeochemical cycles to climate change. Examples from terrestrial ecosystems. *Front. Ecol. Environ.* **9**, 61–67 (2011).
2. McGill, W. B. & Cole, C. V. Comparative aspects of cycling organic C, N, S and P through soil organic matter. *Geoderma* **26**, 267–286 (1981).
3. Peñuelas, J. *et al.* The human-induced imbalance between C, N and P in Earth's life system. *Glob. Change Biol.* **18**, 3–6 (2012).
4. Schlesinger, W. H. *et al.* Biological feedbacks in global desertification. *Science* **247**, 1043–1048 (1990).
5. Vicente-Serrano, S. M. *et al.* Dryness is accelerating degradation of vulnerable shrublands in semiarid Mediterranean environments. *Ecol. Monogr.* **82**, 407–428 (2012).
6. Austin, A. T. *et al.* Water pulses and biogeochemical cycles in arid and semiarid ecosystems. *Oecologia* **141**, 221–235 (2004).
7. Schwinning, S. & Sala, O. E. Hierarchy of responses to resource pulses in arid and semi-arid ecosystems. *Oecologia* **141**, 211–220 (2004).
8. Whitford, W. G. *Ecology of Desert Systems* (Academic, 2002).
9. Gao, X. J. & Giorgi, F. Increased aridity in the Mediterranean region under greenhouse gas forcing estimated from high resolution simulations with a regional climate model. *Global Planet. Change* **62**, 195–209 (2008).
10. Feng, S. & Fu, Q. Expansion of global drylands under a warming climate. *Atmos. Chem. Phys. Discuss.* **13**, 14637–14665 (2013).
11. Dai, A. Increasing drought under global warming in observations and models. *Nature Clim. Change* **3**, 52–58 (2013).
12. Schlesinger, W. H. *Biogeochemistry, an Analysis of Global Change* (Academic, 1996).
13. Walker, T. W. & Syers, J. K. The fate of phosphorus during pedogenesis. *Geoderma* **15**, 1–19 (1976).
14. Vitousek, P. M. *Nutrient Cycling and Limitation: Hawai'i as a Model System* (Princeton Univ. Press, 2004).
15. Nannipieri, P. *et al.* *Phosphorus in Action* (Soil Biol. 26, Springer, 2011).
16. Liebig, J. *et al.* *Chemistry in its Application to Agriculture and Physiology* 3rd edn (Owen, 1842).
17. Reynolds, J. F. *et al.* Global desertification: building a science for dryland development. *Science* **316**, 847–851 (2007).
18. Schimel, D. S. Drylands in the Earth system. *Science* **327**, 418–419 (2010).
19. Maestre, F. T. *et al.* It's getting hotter in here: determining and projecting the impacts of global environmental change on drylands. *Phil. Trans. R. Soc. B* **367**, 3062–3075 (2012).
20. Cross, A. F. & Schlesinger, W. H. Biological and geochemical controls on phosphorus fractions in semiarid soils. *Biogeochemistry* **52**, 155–172 (2001).
21. Li, J. *et al.* Quantitative effects of vegetation cover on wind erosion and soil nutrient loss in a desert grassland of southern New Mexico, USA. *Biogeochemistry* **85**, 317–332 (2007).
22. Sinsabaugh, R. L. *et al.* *Enzymes in the Environment* (Oxford Univ. Press, 2002).
23. Olander, L. P. & Vitousek, P. M. Regulation of soil phosphatase and chitinase activity by N and P availability. *Biogeochemistry* **49**, 175–191 (2000).
24. Schimel, J. P. & Bennett, J. Nitrogen mineralization, challenges of a changing paradigm. *Ecology* **85**, 591–602 (2004).
25. Evans, S. E. & Burke, I. C. carbon and nitrogen decoupling under an 11-year drought in the shortgrass steppe. *Ecosystems (N. Y.)* **16**, 20–33 (2013).

26. Thornton, P. E. *et al.* Influence of carbon–nitrogen cycle coupling on land model response to CO₂ fertilization and climate variability. *Glob. Biogeochem. Cycles* **21**, GB4018 (2007).
27. Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
28. Maestre, F. T. *et al.* Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335**, 214–218 (2012).
29. Hijmans, R. J. *et al.* Very high resolution interpolated climate surfaces for global areas. *Int. J. Clim.* **25**, 1965–1978 (2005).
30. Grace, J. B. *Structural Equation Modelling Natural Systems* (Cambridge Univ. Press, 2006).

Acknowledgements We thank M. Scheffer, N. J. Gotelli and R. Bardgett for comments on previous versions of the manuscript, and all the technicians and colleagues who helped with the field surveys and laboratory analyses. This research is supported by the European Research Council (ERC) under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC Grant agreement no. 242658

(BIOCOM), and by the Ministry of Science and Innovation of the Spanish Government, grant no. CGL2010-21381. CYTED funded networking activities (EPES, Acción 407AC0323). M.D.-B. was supported by a PhD fellowship from the Pablo de Olavide University.

Author Contributions F.T.M., M.D.-B. and A.G. designed this study. F.T.M. coordinated all field and laboratory operations. Field data were collected by all authors except A.E., A.G., B.G., E.V., M.B. and M.D.W. Laboratory analyses were done by V.O., A.G., M.B., M.D.-B., E.V. and B.G. Data analyses were done by M.D.-B. and M.A.B. The paper was written by M.D.-B., F.T.M., M.D.W. and A.G., and the remaining authors contributed to the subsequent drafts.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D.-B. (mdelbaq@upo.es).

METHODS

Study site and data collection. Field data were collected from 224 dryland sites located in 16 countries from all continents except Antarctica (see ref. 28 for full details on the study sites sampled). Locations for this study were chosen to represent a wide spectrum of abiotic (climatic, soil type, slope) and biotic (type of vegetation, total cover, species richness) features characterizing drylands worldwide. At each site, we established a 30 m × 30 m plot representative of the dominant vegetation. Within each plot, we measured plant cover using the line-intercept method along four 30-m-long transects separated 8 m from each other²⁸. Soils were sampled using a stratified random procedure. At each plot, five 50 cm × 50 cm quadrats were randomly placed under the canopy of the dominant perennial plant species and in open areas devoid of perennial vegetation, and a composite sample (0–7.5-cm depth) was obtained from each of them (10–15 soil samples per site were collected; over 2,600 samples were collected and analysed in total). After field collection, soil samples were taken to the laboratory, where they were sieved, air-dried for one month and stored for laboratory analyses. The clay percentage was determined as described in ref. 28.

Selection of soil C, N and P surrogates. The biogeochemical cycles of C, N and P are interlinked with primary production, respiration and decomposition in terrestrial ecosystems^{1,12,31}. All these nutrients have a large variety of forms in soils, including different qualities (labile and recalcitrant) and chemistries^{12,32,33} (organic and inorganic). Only some of them, however, are available for plants and microorganisms^{12,24}. Because of the importance of the biogeochemical cycles of C, N and P on the plant and microorganism stoichiometries^{1,3,34}, we focused our study on the available nutrient forms for these organisms. Thus, we selected total organic C as our surrogate of the C cycle because we observed that it is a parsimonious summary of labile C sources available to soil microorganisms, as this variable is strongly correlated with the availability of other C sources such as dissolved carbohydrates (Pearson's $r = 0.59$; $P < 0.001$) at the studied sites. Similarly, total N was selected as our N-cycle surrogate because of its relationship to other N forms available for plants and microorganisms, such as dissolved inorganic N (Pearson's $r = 0.40$, $P < 0.001$) and amino acids^{12,24,31,33,35} (Pearson's $r = 0.41$, $P < 0.001$). Finally, we selected total P as our P-cycle surrogate because of its relationship to other P forms available for plants and microorganisms^{2,13–15,22,36–41}. In our study sites, total P was positively related to available P (Olsen inorganic P; Pearson's $r = 0.43$, $P < 0.001$) and HCl-P (Pearson's $r = 0.23$, $P = 0.001$). The Olsen inorganic P extract (available P) simulates the action of plant roots in dissolving P minerals, and is considered as an index of plant-available P, and the 1 M HCl inorganic P fraction (non-occluded) represents the P associated with calcium carbonate minerals, which is available over long time periods^{20,37–39,42,43}. The exchange between P and carbonate minerals limits P availability in desert soils, and the precipitation of phosphate with calcium establishes an upper limit for the availability of P (refs 5,22,42–44).

To avoid problems associated with the use of multiple laboratories when analysing soils from different sites, and to facilitate the comparison of results between them, dried soil samples from all the countries were shipped to Spain for analyses. All the analyses for organic C, Olsen inorganic P, total P and phosphatase activity were carried out at the laboratory of the Biology and Geology Department, Rey Juan Carlos University (Móstoles, Spain). Analyses of total N were carried out at the University of Jaén (Jaén, Spain). The remaining soil analyses were carried out at the laboratory of the Department of Physical, Natural and Natural Systems, Pablo de Olavide University (Sevilla, Spain). Organic C was determined by colorimetry after oxidation with a mixture of potassium dichromate and sulphuric acid⁴⁵. Total N was obtained using a CN analyser (Leco CHN628 Series, Leco Corporation). Total P was measured using a Skalar San++ Analyzer after digestion with sulphuric acid. Olsen inorganic P was measured following a 0.5 M NaHCO₃ (pH 8.5) extraction^{36,37}. Soil extracts in a ratio of 1:5 were shaken in a reciprocal shaker at 200 r.p.m. for 2 h. An aliquot of the centrifuged extract was used for the colorimetric determination of Olsen inorganic P, on the basis of its reaction with ammonium molybdate⁴⁶; the pH of the extracts was adjusted with 0.1 M HCl when necessary. The 1 M P HCl-P fraction was determined following the soil-P fractionation protocol of ref. 38. The remaining soil variables were measured in K₂SO₄ 0.5 M soil extracts in the ratio 1:5. Soil extracts were shaken on an orbital shaker at 200 r.p.m. for 2 h at 20 °C and filtered to pass a 0.45-µm Millipore filter. The filtered extract was kept at 2 °C until colorimetric analyses, which were conducted within 24 h of the extraction. Subsamples of each extract were taken for measurements of carbohydrates (sum of hexoses and pentoses) and amino acids according to ref. 35. Inorganic-N (sum of ammonium and nitrate) concentrations were also measured for each K₂SO₄ 0.5 M extract subsample with the indophenol blue method described in ref. 47.

Rationale of the variables included in structural equation modelling. Aridity is a fundamental driver of biological and geochemical processes^{5,8,12,48–50} in arid, semi-arid, and dry sub-humid ecosystems (areas where the ratio of precipitation

to potential evapotranspiration is less than 0.65 (ref. 28); hereafter drylands), where water availability is the most limiting resource^{6–8,17,18}. Aridity determines water availability in drylands, and therefore has a substantial impact on factors such as plant productivity, microbial activity, nutrient concentration and soil enzyme activities^{6–8,12,17,18}. This environmental factor has both direct and indirect impacts on ecosystem services, and on multiple processes directly related to ecosystem functioning^{5,8}. For example, increasing aridity in drylands has been observed to decrease vegetation cover⁴³, indirectly promoting soil erosion by wind^{3–5,8,17,18,48,51}, which can subsequently lead to land degradation and desertification^{3–5,8,17,18,48,51}. Wind erosion can remove silt, clay and organic matter from the surface soil, leaving behind sand and infertile materials⁵¹. In addition, aridity promotes soil drying, increasing its salinity levels and enhancing soil erosion, which effects remove fine, nutrient-rich particles such as clay^{3–5,8,17,18,20,43,51}.

The cover of perennial vascular plants is also a key driver of ecosystem structure and functioning in drylands, because this variable largely determines processes such as plant facilitation, litter production and decomposition, and biological N fixation, as well as the ability of landscapes to retain water and nutrients^{52–58}. Therefore, plant cover is closely related to nutrient availability in dryland soils^{12,53–58}.

Clay has an important role on the retention of water and nutrients at the soil surface, where microbial activity is greatest, and can also modify local pH^{59–63}. The activity of phosphatase was included in our structural equation model because extracellular enzymes are proximate agents of organic matter decomposition and their assessment can be used as an indicator of microbial nutrient demand^{15,22}. The activity of extracellular enzymes, which are produced by both plants and microorganisms^{15,22}, is known to be negatively affected by factors linked to aridity such as low water availability and soil salinity^{15,22,64}. Enzyme activities have been observed to be associated with clay abundance in soil^{63,64}. Phosphatase is the enzyme responsible for releasing inorganic P from organic sources, and is considered a surrogate of biological P demand^{45,22}. Phosphatase activity was measured by determination of the amount of *p*-nitrophenol released from 0.5 g soil after incubation at 37 °C for 1 h with the substrate *p*-nitrophenyl phosphate in MUB buffer (pH 6.5). Inorganic P is a universal source of P for plants and microorganisms^{15,20,22,44}. As a component of biological molecules fundamental to cellular energy transfer (that is, ATP and nicotinamide adenine trinucleotide), P has a major role in the C and N fixation in drylands, and is non-biological in origin, being derived from rocks and sediments¹. Carbon and N are primarily linked to biological processes such as photosynthesis, atmospheric-N fixation and subsequent microbial mineralization¹², and are considered to be key elements for enzyme production^{12,14,22,23}. Finally, we included the spatial location (distance from Equator and longitude) of each site in our structural equation model to account for the spatial autocorrelation present in our data (see ref. 28 for a related approach).

Statistical and numerical analyses. Before numerical and statistical analysis, all the variables in this study (plant cover, clay, organic C, total N, total P and activity of phosphatase) were averaged to obtain site-level estimates by using the mean values observed in bare ground and vegetated areas, weighted by their respective cover at each site.

We explored the relationship of aridity with the different selected C (organic C), N (total N) and P (total P) variables, and with the N:P, C:P and N:C ratios, by using either linear or curvilinear (quadratic) regressions. Similarly, we explored the relationship between aridity and inorganic P (sum of Olsen inorganic P and HCl-P), organic C and total N, and that with their respective N:P, C:P and N:C ratios (Extended Data Fig. 1). Among these, the function that minimized the second-order Akaike information criterion⁶⁵ was chosen in each case. All the nutrient ratios were log-transformed to achieve normality before conducting these analyses. The aridity index^{8,19,51} (AI, the ratio of precipitation to potential evapotranspiration) of each site was calculated using data interpolations provided by WorldClim^{29,66}. To facilitate the interpretation of our results, we used $1 - AI$ as our surrogate of aridity. This index increases with decreasing annual mean precipitation in our database ($r = 0.91$, $P < 0.001$). We also explored the relationship of aridity with more labile sources of C (carbohydrates), N (sum of dissolved inorganic N and amino acids) and available P (Olsen inorganic P), as well as with their respective ratios. The results obtained (Extended Data Fig. 2) were very similar to those presented in the main text (Fig. 1); hence, we used total N, organic C and total P there because of their typically high stability in time^{67–69}.

To determine the relative importance of aridity on the selected soil nutrients differentially linked to biological (C and N) versus geochemical (P) control, we used structural equation modelling³⁰ (SEM). Overall, SEM has emerged as a synthesis of path analysis, factor analysis and maximum-likelihood techniques, and has been thoroughly used in the ecological sciences as a causal inference tool^{30,70}. It can test the plausibility of a causal model, which is based on a-priori information on the relationships among the particular variables of interest. Some data manipulation was required before modelling. We checked the bivariate relationships between all variables to ensure that a linear model was appropriate. We identified

some curvilinear relationships among our variables. Several variables showed a curvilinear relationship with latitude, such that areas closer to the Equator tended to be different from areas farther from the Equator. This was simply handled by expressing latitude as distance from the Equator (that is, the absolute value of latitude). Because longitude has an arbitrary origin, this transformation did not apply to longitude. We found that organic C, total N and the activity of phosphatase were curvilinearly influenced by aridity, and that these relationships were well described by a second-order polynomial. To introduce polynomial relationships into our model, we calculated the square of aridity and introduced it into our model using a composite variable approach described below. We also examined the distributions of all of our endogenous variables, and tested their normality. Organic C, total N, total P and activity of phosphatase were log-transformed to improve normality. Similarly, the plant total cover was square-root-transformed. Because organic C and total N were very closely related (Pearson's $r = 0.92$), we could not introduce them into the same model without risking collinearity. Attempts to construct a latent variable including organic C and total N were not successful. Thus, we reduced these two variables to a single variable using PCA on the correlation matrix, and then introduced this new variable into the model (organic-matter component). We interpret this variable as organic C and N, as both variables were very highly correlated with the PCA component (Pearson's $r = 0.98$); although some of the total N is certainly in mineral form⁷¹, this very close relationship indicates that total N is under tight control of organic matter in the studied drylands. Then, we established an a-priori model (Extended Data Fig. 4) based on the known effects and relationships among the drivers of C, N and P availability. This model accounted for spatial structure (latitude and longitude), aridity, percentage of plant cover and clay, organic matter component (total N and organic C), activity of phosphatase and total P.

When these data manipulations were complete, we parameterized our model using our data set and tested its overall goodness of fit. There is no single universally accepted test of overall goodness of fit for structural equation models, applicable in all situations regardless of sample size or data distribution. Most modellers circumvent this problem by using multiple goodness-of-fit criteria. We used the χ^2 -test (the model has a good fit when $0 \leq \chi^2 \leq 2$ and $0.05 < P \leq 1.00$) and the root mean square error of approximation (RMSEA; the model has a good fit when $0 \leq \text{RMSEA} \leq 0.05$ and $0.10 < P \leq 1.00$). Additionally, and because some variables were not normally distributed, we confirmed the fit of the model using the Bollen–Stine bootstrap test⁷² (the model has a good fit when $0.10 < \text{bootstrap } P \leq 1.00$). Our a-priori model attained an acceptable fit by all criteria, and thus no *post hoc* alterations were made.

After attaining a satisfactory model fit, we introduced composite variables into our model. The use of composite variables does not alter the underlying SEM model, but collapses the effects of multiple conceptually related variables into a single composite effect, aiding interpretation of model results^{30,70}. Distance from the Equator and longitude were included as a composite variable, because together they determine the spatial proximity of plots. A separate composite was constructed for each response variable. We also used composite variables to model the nonlinear response to aridity of the organic matter component (first component from a PCA with organic C and total N) and phosphatase activity. As previously mentioned, both aridity and its square are introduced as variables in the model. Because one is mathematically derived from the other, they are allowed to covary. In cases where a nonlinear fit is desired, the effects of aridity and squared aridity on a given response are composited. The resulting effect has no interpretable sign, because the relationship may be positive over some portion of the data and negative over other portions. In cases where a simple linear effect of aridity was desired (for example total P), we simply included a single path from aridity and did not use squared aridity.

With a reasonable model fit, and composite variables constructed, we were free to interpret the path coefficients of the model, and their associated *P* values. A path coefficient is analogous to a partial correlation coefficient, and describes the strength and sign of the relationships between two variables^{30,70}. Because some of the variables introduced were not normally distributed, the probability that a path coefficient differs from zero was tested using bootstrap tests^{30,70}. Bootstrapping is preferred to the classical maximum-likelihood estimation in these cases, because in bootstrapping probability assessments are not based on an assumption that the data match a particular theoretical distribution. Thus, data are randomly sampled with replacement to arrive at estimates of standard errors that are empirically associated with the distribution of the data in the sample^{30,70}.

Another important capability of SEM is its ability to partition direct and indirect effects that one variable may have on another, and estimate the strengths of these multiple effects. Unlike regression or analysis of variance, SEM offers the ability to separate multiple pathways of influence and view them as a system^{30,70}. Thus, SEM is useful for investigating the complex networks of relationships found in natural ecosystems^{30,70}.

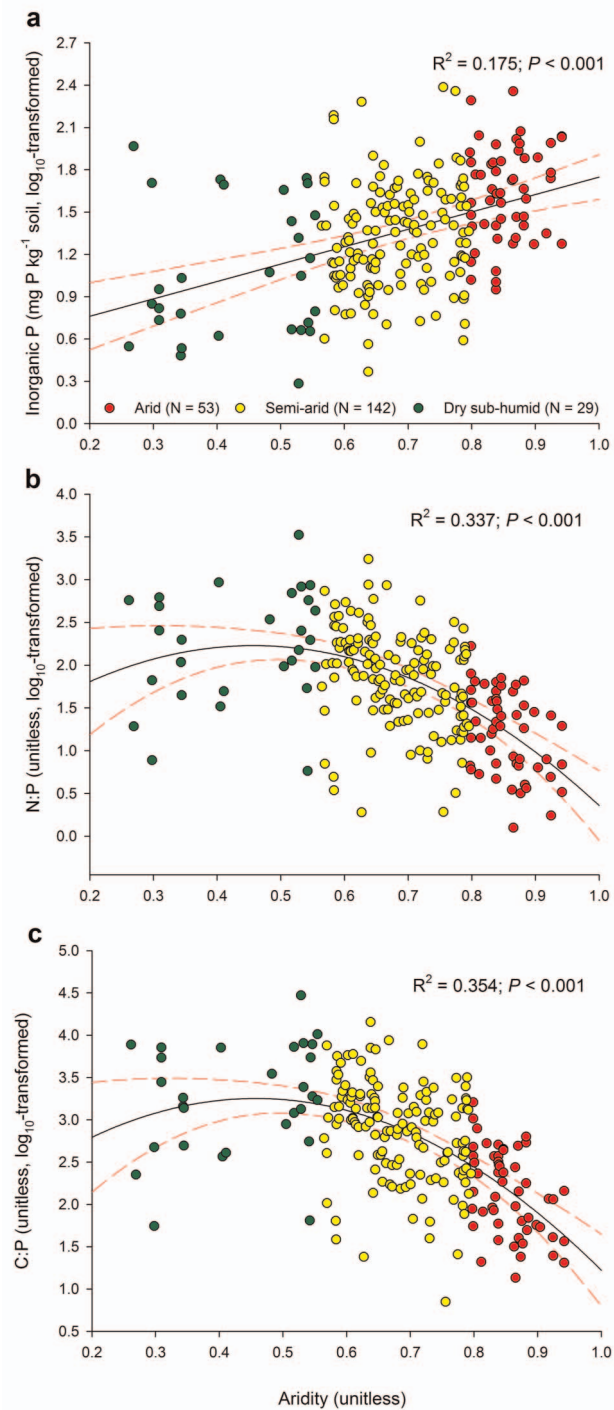
To aid final interpretation in light of this ability of SEM, we calculated not only the standardized total effects of aridity, percentage of clay and plant cover, and spatial position (distance from Equator and longitude) on the organic matter component, but also the effect of the organic matter component on phosphatase and total P activity. The net influence that one variable has on another is calculated by summing all direct and indirect pathways between the two variables. If the model fits the data well, the total effect should approximate the bivariate correlation coefficient for that pair of variables^{30,70}.

In addition, and to support our results further, we repeated our structural equation model including inorganic P (sum of Olsen inorganic P and HCl-P) instead of total P (Extended Data Figs 5 and 6) and using a labile organic matter component (first component from a PCA with carbohydrates and available N) and available P (Olsen inorganic P; Extended Data Figs 7 and 8).

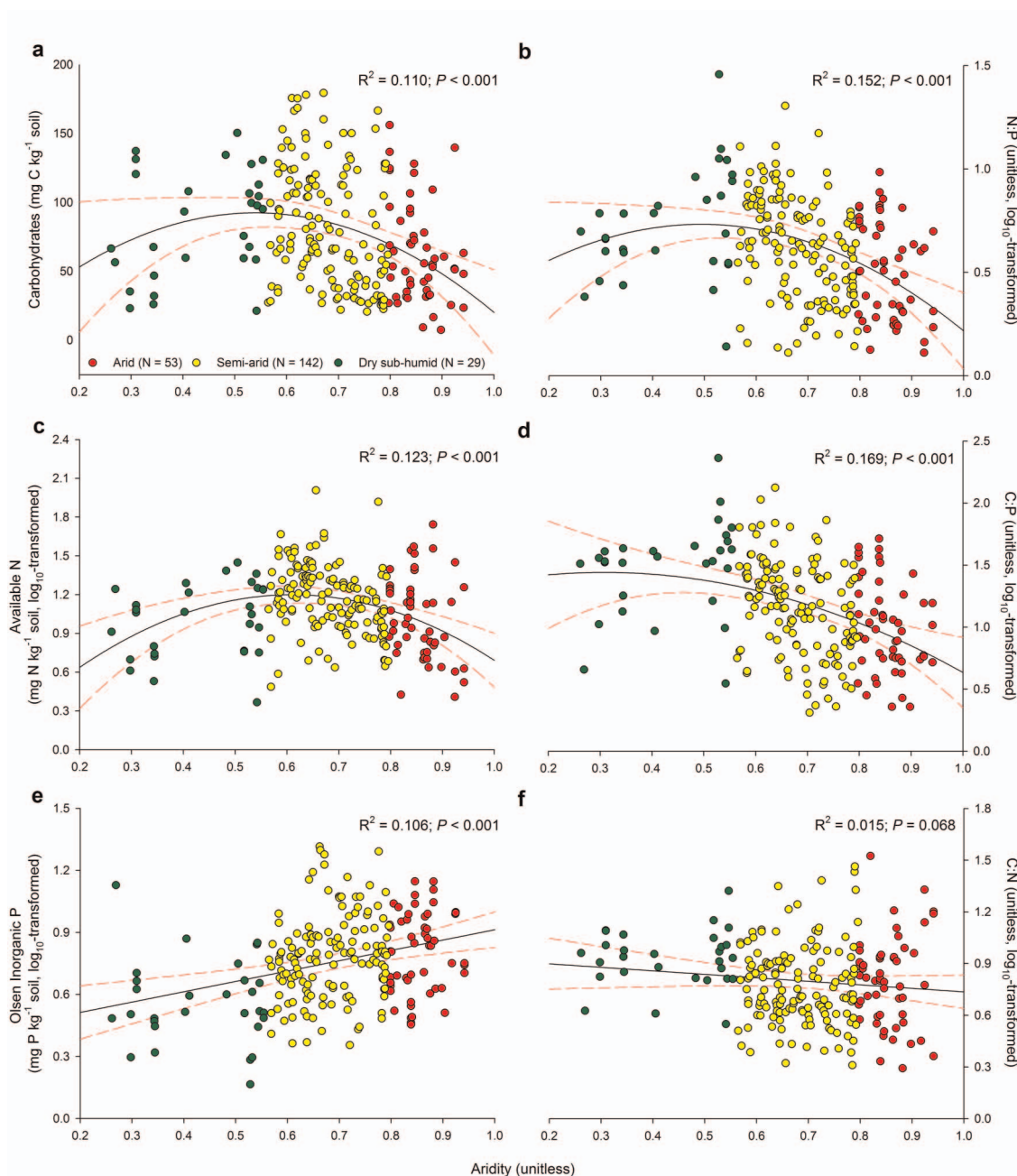
All the SEM analyses were conducted using AMOS 18.0 (Amos Development Co.). The remaining statistical analyses were performed using SPSS 15.0 (SPSS Inc.).

31. Robertson, G. P. & Groffman, P. *Soil Microbiology and Biochemistry* (Springer, 2007).
32. Rovira, P. & Vallejo, V. R. Labile and recalcitrant pools of carbon and nitrogen in organic matter decomposing at different depths in soil: an acid hydrolysis approach. *Geoderma* **107**, 109–141 (2002).
33. Neff, J. C. *et al.* Breaks in the cycle: dissolved organic nitrogen in terrestrial ecosystems. *Front. Ecol. Environ* **1**, 205–211 (2003).
34. Sardans, J. *et al.* The C:N:P stoichiometry of organisms and ecosystems in a changing world: a review and perspectives. *Perspect. Plant Ecol. Evol. Syst.* **14**, 33–47 (2012).
35. Chantigny, M. H. *et al.* *Soil Sampling and Methods of Analysis* (CRC, 2006).
36. Bray, R. H. & Kurtz, L. T. Determination of total, organic, and available forms of phosphorus in soils. *Soil Sci.* **59**, 39–46 (1945).
37. Olsen, S. R. *et al.* Estimation of available phosphorus in soils by extraction with sodium bicarbonate. *USDA Circ.* **939** (1954).
38. Tiessen, H. & Moir, J. O. *Characterization of Available P by Sequential Fractionation. Soil Sampling and Methods of Analysis* (Lewis, 1993).
39. Carreira, J. A. *et al.* Phosphorus transformations along a soil/vegetation series of fire-prone, dolomitic, semi-arid shrublands of southern Spain. *Biogeochemistry* **39**, 87–120 (1997).
40. Schoenau, J. J. *et al.* Forms and cycling of phosphorus in prairie and boreal forest soils. *Biogeochemistry* **8**, 223–237 (1989).
41. Bowman, R. A. & Cole, C. V. Transformations of organic phosphorus substrates in soils as evaluated by NaHCO₃ extractions. *Soil Sci.* **125**, 49–54 (1978).
42. Cross, A. F. & Schlesinger, W. H. A literature review and evaluation of the Hedley fractionation: applications to the biogeochemical cycle of soil phosphorus in natural ecosystems. *Geoderma* **64**, 197–214 (1995).
43. Lajtha, K. & Bloomer, S. H. Factors affecting phosphate sorption and phosphate retention in a desert ecosystem. *Soil Sci.* **146**, 160–167 (1988).
44. Roberts, T. L. *et al.* The influence of topography on the distribution of organic and inorganic soil phosphorus across a narrow environmental gradient. *Can. J. Soil Sci.* **65**, 651–665 (1985).
45. Anderson, J. M. & Ingram, J. S. I. *Tropical Soil Biology and Fertility: A Handbook of Methods* 2nd edn (CABI, 1993).
46. Tabatabai, M. A. & Bremner, J. M. Use of p-nitrophenyl phosphate for assay of soil phosphatase activity. *Soil Biol. Biochem.* **1**, 301–307 (1969).
47. Delgado-Baquerizo, M. *et al.* A dissolved organic nitrogen in Mediterranean ecosystems. *Pedosphere* **21**, 309–318 (2011).
48. Helmut, G. *The Causes and Progression of Desertification* (Ashgate, 2005).
49. Michael, B. *et al.* *Human Impact on Environment and Sustainable Development in Africa* (Ashgate, 2003).
50. Johnson, P. J. *Governing Global Desertification: Linking Environmental Degradation, Poverty and Participation* (Ashgate, 2006).
51. Food and Agriculture Organization. *Arid Zone Forestry: A Guide for Field Technicians* Ch. I (Food and Agriculture Organization, 1989).
52. Vitousek, P. M. *et al.* Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry* **57**, 1–45 (2002).
53. Maestre, F. T. *et al.* Potential of using facilitation by grasses to establish shrubs on a semiarid degraded steppe. *Ecol. Appl.* **11**, 1641–1655 (2001).
54. Reynolds, J. F. *et al.* Impact of drought on desert shrubs: effects of seasonality and degree of resource island development. *Ecol. Monogr.* **69**, 69–106 (1999).
55. Maestre, F. T. *et al.* Positive, negative and net effects in grass-shrub interactions in Mediterranean semiarid grasslands. *Ecology* **84**, 3186–3197 (2003).
56. Eldridge, D. *et al.* Interactive effects of three ecosystem engineers on infiltration in a semi-arid Mediterranean grassland. *Ecosystems* (N. Y.) **13**, 499–510 (2010).
57. Cerdà, A. The effect of patchy distribution of *Stipa tenacissima* L. on runoff and erosion. *J. Arid Environ.* **36**, 37–51 (1997).
58. Cornelis, W. S. *Dryland Ecohydrology* (Springer, 2006).
59. Blanco, H. & Rattan, R. *Principles of Soil Conservation and Management* (Springer, 2010).
60. Yerima, P. K. *et al.* *Introduction to Soil Science: Soils of the Tropics* (Trafford, 2005).
61. Marshall, K. C. Clay mineralogy in relation to survival of soil bacteria. *Annu. Rev. Phytopathol.* **13**, 357–373 (1975).
62. Stotzky, G. Activity, ecology, and population dynamics of microorganisms in soil. *Rev. Microbiol.* **2**, 59–137 (1972).
63. Kandeler, E. *Physiological and Biochemical Methods for Studying Soil Biota and Their Function. Soil Microbiology and Biochemistry* (Springer, 2007).

64. Tietjen, T. & Wetzel, R. G. Extracellular enzyme-clay mineral complexes: enzyme adsorption, alteration of enzyme activity, and protection from photodegradation. *Aquat. Ecol.* **37**, 331–339 (2003).
65. Sugiura, N. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Stat. Theor. M.* **A7**, 13–26 (1978).
66. Zomer, R. J. *et al.* *Carbon, Land and Water: a Global Analysis of the Hydrologic Dimensions of Climate Change Mitigation Through Afforestation/Reforestation*. Research Report 101 (International Water Management Institute, 2006).
67. Schimel, D. S. *et al.* Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Glob. Biogeochem. Cycles* **8**, 279–293 (1994).
68. Schmidt, M. W. I. *et al.* Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).
69. Oades, J. M. The retention of organic matter in soils. *Biogeochemistry* **5**, 35–70 (1988).
70. Shipley, B. *Cause and Correlation in Biology: a User's Guide to Path Analysis Structural Equations and Causal Inference* (Cambridge Univ. Press, 2001).
71. Morford, S. L. *et al.* Increased forest ecosystem carbon and nitrogen storage from nitrogen rich bedrock. *Nature* **477**, 78–81 (2011).
72. Schermelleh-Engel. *et al.* Evaluating the fit of structural equation models, tests of significance descriptive goodness-of-fit measures. *Methods Psychol. Res. Online* **8**, 23–74 (2003).

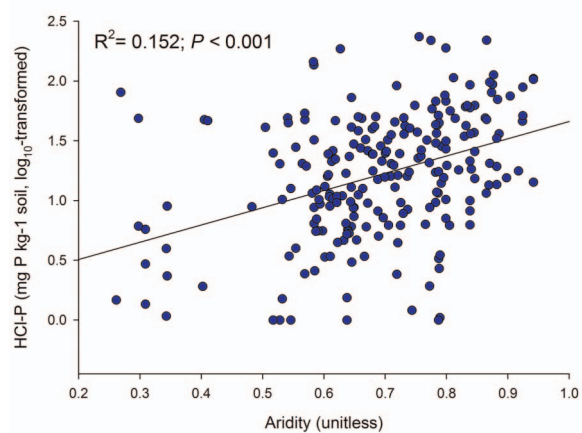


Extended Data Figure 1 | Relationships between aridity and the concentration of inorganic P and the ratios of total N to inorganic P and organic C to inorganic P at our study sites. Inorganic P, sum of Olsen inorganic P and HCl-P. The solid and dashed lines represent the fitted quadratic regressions and their 95% confidence intervals, respectively.

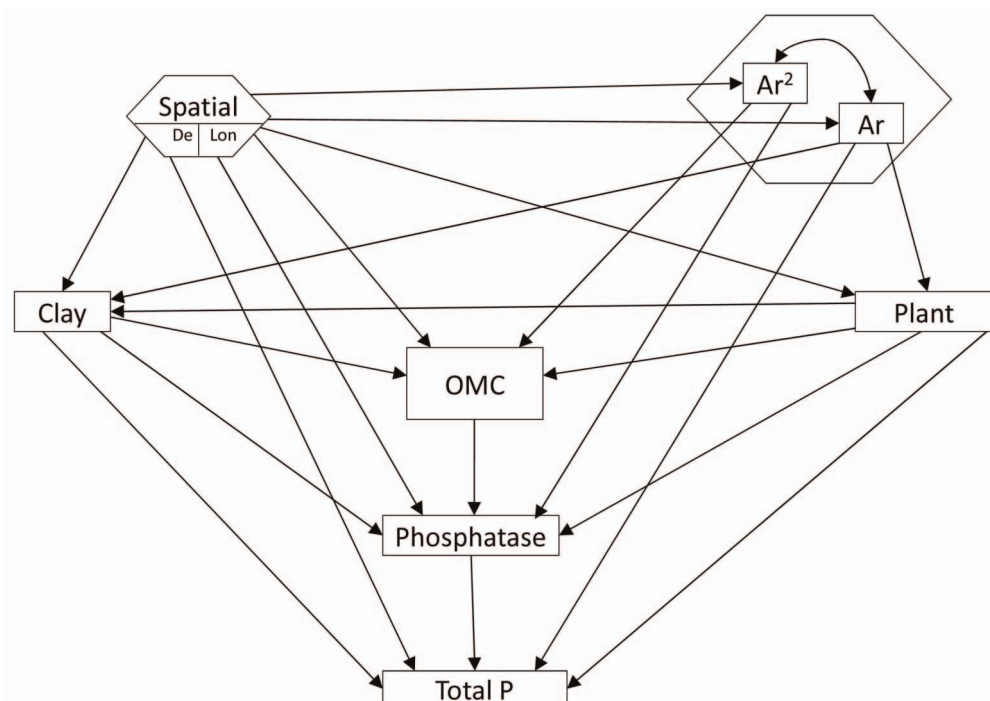


Extended Data Figure 2 | Relationships between aridity and the concentration of carbohydrates (C), available N, available P and their ratios at our study sites. Available N, sum of dissolved inorganic N and amino acids;

available P, Olsen inorganic P. The solid and dashed lines represent the fitted quadratic regressions and their 95% confidence intervals, respectively.

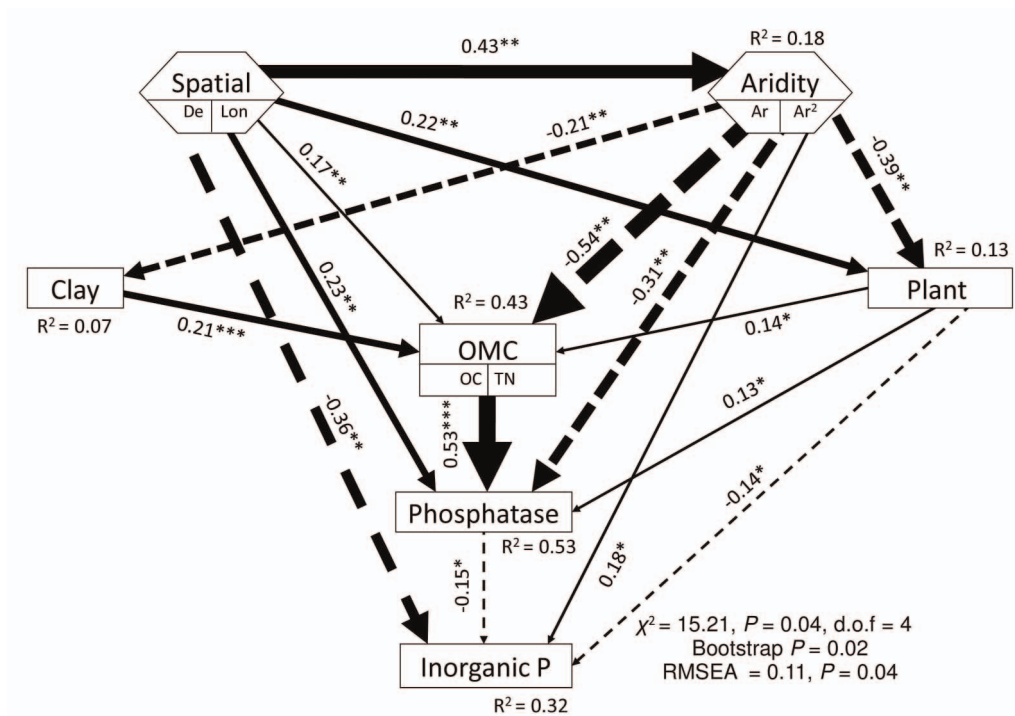


Extended Data Figure 3 | Relationships between aridity and the concentration of HCl-P fraction at our study sites.



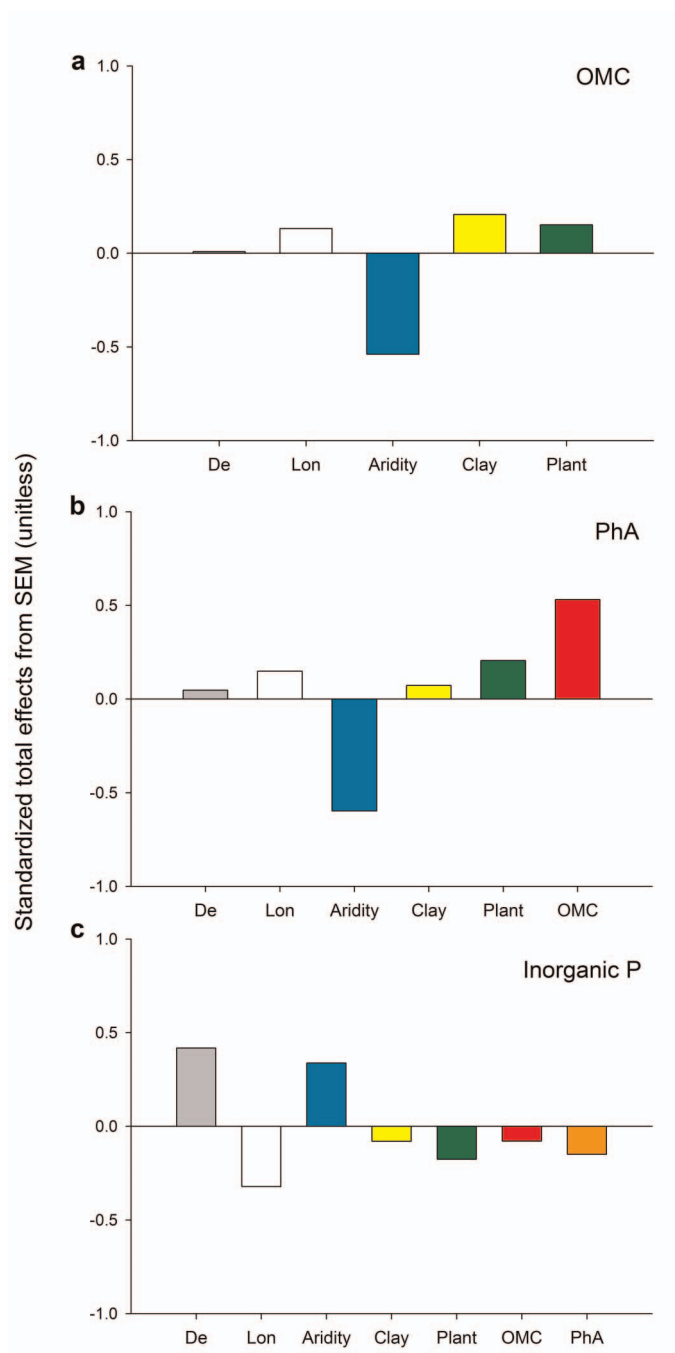
Extended Data Figure 4 | A-priori structural equation model used in this study. We included in this model aridity (*Ar*; composite variable formed from *Ar* and *Ar*²), percentage of plant cover (*Plant*), percentage of clay (*Clay*), spatial position (*Spatial*; composite variable formed from distance from Equator (*De*) and longitude (*Lon*)), activity of phosphatase, organic matter component (*OMC*; first component from a PCA conducted with organic C

(*OC*) and total N (*TN*)) and total P. We built our structural equation model by taking into account all these relationships, as explained in Methods. There are some differences between the a-priori model and the final model structures owing to removal of paths with coefficients close to zero (Fig. 2). Hexagons are composite variables³⁰. Squares are observable variables.

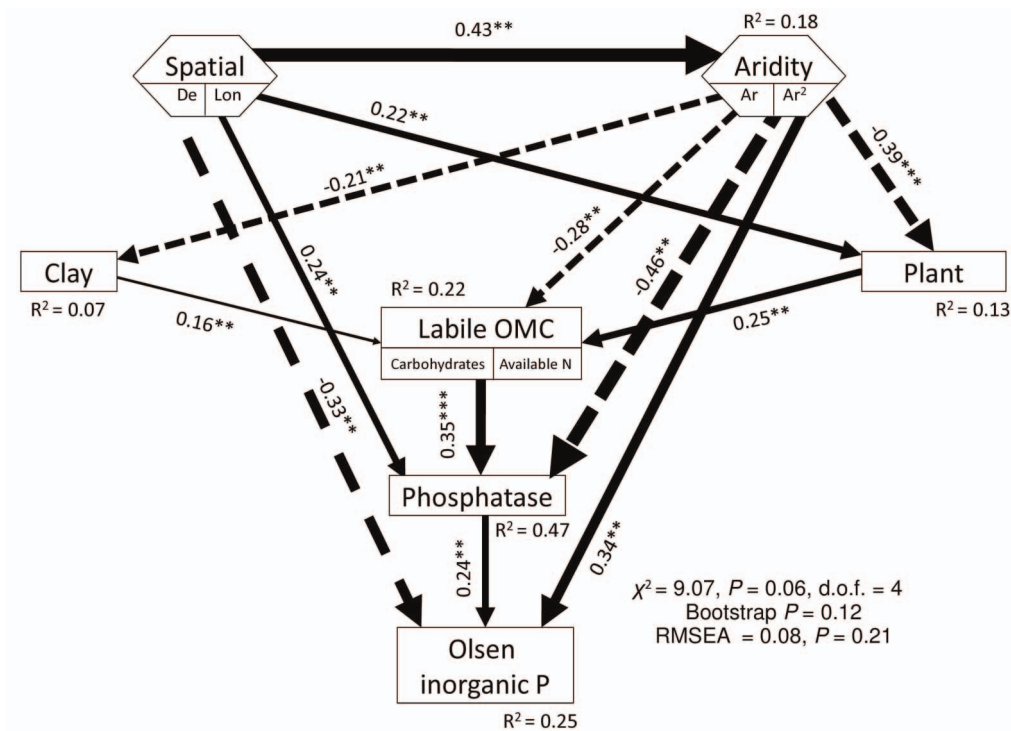


Extended Data Figure 5 | Global structural equation model, depicting the effects of aridity, clay percentage, plant cover and site position on the organic matter component, the inorganic-P concentration and phosphatase activity. Spatial coordinates of the study sites are expressed in terms of distance from Equator (De) and longitude (Lon). The organic matter component (OMC) is the first component from a PCA conducted with organic C and total N. The inorganic-P concentration is the sum of Olsen inorganic P and HCl-P. Numbers adjacent to arrows are standardized path coefficients, analogous to relative regression weights, and indicative of the effect size of the relationship.

Continuous and dashed arrows indicate positive and negative relationships, respectively. The width of arrows is proportional to the strength of path coefficients. The proportion of variance explained (R^2) appears above every response variable in the model. Goodness-of-fit statistics for each model are shown in the lower right corner. There are some differences between the a-priori model and the final model structures owing to removal of paths with coefficients close to zero (see the a-priori model in Extended Data Fig. 4). Hexagons are composite variables³⁰. Squares are observable variables. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

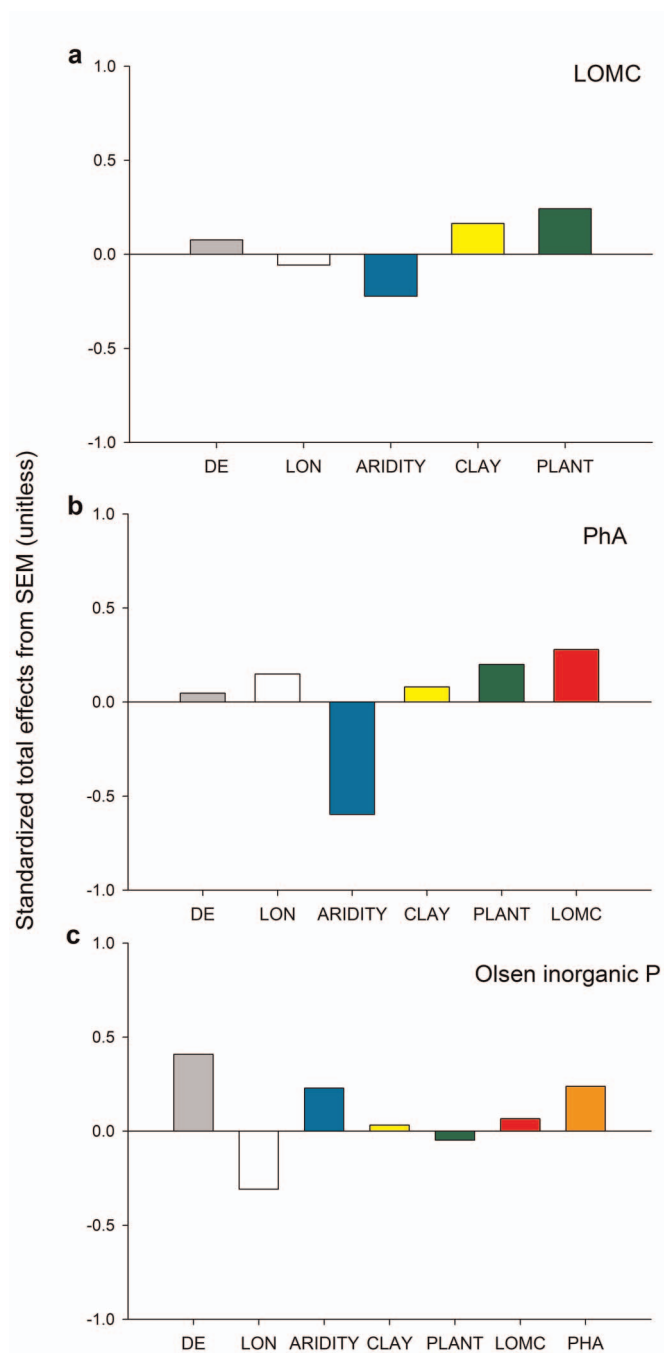


Extended Data Figure 6 | Standardized total effects (direct plus indirect effects) derived from the structural equation modelling. These include the effects of aridity, percentage of clay, plant cover, distance from Equator (De) and longitude (Lon) on the organic matter component (OMC, first component from a PCA conducted with organic C and total N), inorganic P (sum of Olsen inorganic P and HCl-P) and phosphatase activity (PhA).

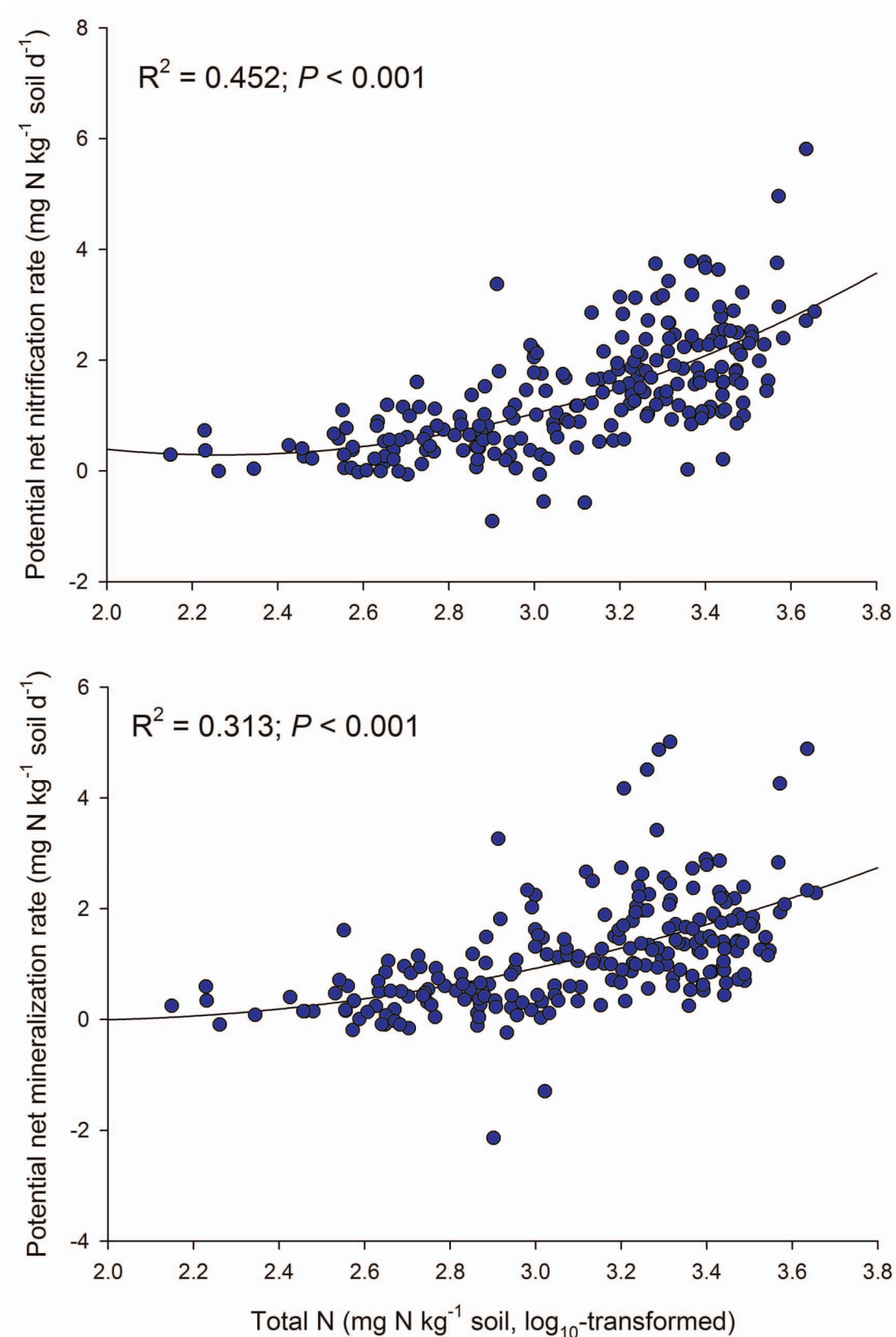


Extended Data Figure 7 | Global structural equation model, depicting the effects of aridity, clay percentage, plant cover and site position on the labile organic matter component, available-P concentration and phosphatase activity. The labile organic matter component (labile OMC) is the first component from a PCA conducted with soil carbohydrates and the ratio of available N to the sum of dissolved inorganic N and amino acids. Available P is the Olsen inorganic P. Numbers adjacent to arrows are standardized path coefficients, analogous to relative regression weights, and indicative of the effect size of the relationship. Continuous and dashed arrows indicate positive and

negative relationships, respectively. The width of arrows is proportional to the strength of path coefficients. The proportion of variance explained (R^2) appears above every response variable in the model. Goodness-of-fit statistics for each model are shown in the lower right corner. There are some differences between the a-priori model and the final model structures owing to removal of paths with coefficients close to zero (see the a-priori model in Extended Data Fig. 4). Hexagons are composite variables³⁰. Squares are observable variables. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

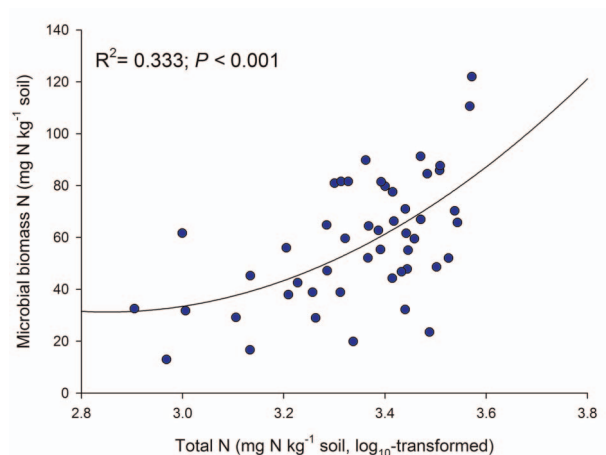


Extended Data Figure 8 | Standardized total effects (direct plus indirect effects) derived from the structural equation modelling. These include the effects of aridity, percentage of clay, plant cover, distance from Equator (De) and longitude (Lon) on the labile organic matter component (LOMC, first component from a PCA conducted with carbohydrates and available N), available P (Olsen inorganic P) and phosphatase activity (PhA).



Extended Data Figure 9 | Relationships between total N and the potential net nitrification (upper graph) and mineralization rates (lower graph) measured at our study sites. Air-dried soil samples were re-wetted to reach 80% of field water-holding capacity and incubated in the laboratory for 14 days

at 30 °C (ref. 28). Potential net nitrification and ammonification rates were estimated as the difference between initial and final nitrate and ammonium concentrations²⁸. The solid line denotes the quadratic model fitted to the data (R^2 and P values shown in each panel).



Extended Data Figure 10 | Relationships between the total N and microbial biomass N in a subset of 50 of our 224 sites. All air-dried soil samples were adjusted to 55% of their water-holding capacity previous to the analyses of microbial biomass N. Microbial biomass N was determined using the fumigation–extraction method. Non-incubated and incubated soil subsamples were fumigated with chloroform for five days. Non-fumigated replicates were used as controls. Fumigated and non-fumigated samples were extracted with K₂SO₄ 0.5 M in the ratio 1:5 and filtered through a 0.45-μm Millipore filter. Concentration of microbial biomass N was estimated as the difference between total N of fumigated and non-fumigated digested extracts²⁸ and then divided by 0.54 (that is, by Kn, the fraction of biomass N extracted after the CHC₁₃ treatment). The solid line denotes the quadratic model fitted to the data (R^2 and P values shown in the graph).

DMSP biosynthesis by an animal and its role in coral thermal stress response

Jean-Baptiste Raina^{1,2,3}, Dianne M. Tapiolas², Sylvain Forêt^{3,4}, Adrian Lutz^{2,3,5}, David Abrego^{2†}, Janja Ceh⁶, François O. Seneca^{1,2†}, Peta L. Clode^{7,8}, David G. Bourne², Bette L. Willis^{1,3} & Cherie A. Motti²

Globally, reef-building corals are the most prolific producers of dimethylsulphoniopropionate (DMSP)^{1,2}, a central molecule in the marine sulphur cycle and precursor of the climate-active gas dimethylsulphide^{3,4}. At present, DMSP production by corals is attributed entirely to their algal endosymbiont, *Symbiodinium*². Combining chemical, genomic and molecular approaches, we show that coral juveniles produce DMSP in the absence of algal symbionts. DMSP levels increased up to 54% over time in newly settled coral juveniles lacking algal endosymbionts, and further increases, up to 76%, were recorded when juveniles were subjected to thermal stress. We uncovered coral orthologues of two algal genes recently identified in DMSP biosynthesis, strongly indicating that corals possess the enzymatic machinery necessary for DMSP production. Our results overturn the paradigm that photosynthetic organisms are the sole biological source of DMSP, and highlight the double jeopardy represented by worldwide declining coral cover, as the potential to alleviate thermal stress through coral-produced DMSP declines correspondingly.

It is widely accepted that the production of dimethylsulphoniopropionate (DMSP), a central molecule in the marine sulphur cycle, is restricted to marine algae and a few species of intertidal plants⁵. Marine bacteria subsequently use DMSP as a source of sulphur and carbon and can metabolize this compound into the volatile gas dimethylsulphide (DMS)⁶, by which the largest natural flux of sulphur enters the atmosphere where it exerts considerable influence on atmospheric chemistry⁷. Despite recent controversy regarding the impact of DMS emissions on global climate⁸, DMS is probably involved in local climate regulation through its oxidation into aerosol particles that induce the formation of clouds and increase their reflectivity, thereby reducing light levels and water temperatures in the marine environment^{3,4}.

Concentrations of DMSP and DMS found in reef-building corals are among the highest recorded in the environment^{1,2}, but it has been assumed that DMSP production derives solely from the coral's endosymbiotic microalgae *Symbiodinium*. Evidence that the total amounts of DMSP recorded in corals are consistently higher than levels present in *Symbiodinium* cells alone^{9,10} raises the possibility of a cryptic source of DMSP in reef-building corals. A clear understanding of the sources of DMSP on reefs and the possible effects that global warming may have on DMSP production is paramount, given the influence that coral-reef-derived sulphur emissions may have on local climates¹¹. Corals in the genus *Acropora* are the most abundant reef-building organisms in the Indo-Pacific region¹² and, as broadcast spawners, they acquire *Symbiodinium* from their surrounding environment after larval development. The *Symbiodinium*-free larvae of this genus provide a unique opportunity to investigate *Symbiodinium*-independent production of DMSP in corals. Results presented here demonstrate that coral hosts

(kingdom: Animalia) are capable of biosynthesizing DMSP in high concentrations in the absence of symbiotic microalgae, refuting the current paradigm that photosynthetic organisms are the only DMSP sources in the environment.

DMSP levels were quantified in juveniles lacking photosynthetic symbionts for two common coral species, *Acropora millepora* and *A. tenuis*, using nuclear magnetic resonance (NMR) spectroscopy¹³. Juveniles were raised in the dark in 0.2- μ m-filtered seawater to ensure they remained algae-free, and the absence of any photosynthetic organisms in these juveniles was confirmed using five different DNA markers, ranging from *Symbiodinium*-specific to universal 23S rRNA plastids primers, targeting all eukaryotic algae and cyanobacteria. A lack of any detectable amplification was observed for all markers (Extended Data Tables 1 and 5). Despite the complete absence of any photosynthetic microalgae, high concentrations of DMSP, two orders of magnitude greater than concentrations reported for benthic algae from the Great Barrier Reef², were recorded in all coral juvenile samples (Fig. 1). Repeated sampling over a 6-day period after larval settlement revealed that the initially high DMSP concentrations in juveniles increased significantly over time; specifically, DMSP increased by 44% (that is, 1.1 nmol mm⁻²) in *A. millepora* and by 54% (that is, 1.7 nmol mm⁻²) in *A. tenuis* (ANOVA, $n = 6$, $P < 0.005$; Fig. 1 and Extended Data Table 2). These results demonstrate unambiguously that the DMSP levels measured were not inherited from parent colonies but instead were produced by coral juveniles growing in the absence of photosynthetic symbionts.

Increases in DMSP concentrations over time were even more pronounced when coral juveniles experienced thermal stress (32 °C), with *A. tenuis* and *A. millepora* exhibiting 65% and 76% increases over the 6-day experimental period, respectively (ANOVA, $n = 6$, $P < 0.005$; Fig. 1 and Extended Data Table 2). Conversely, concentrations of the DMSP breakdown product acrylate decreased in both species, with thermally stressed juveniles containing 33% (*A. millepora*) and 61% (*A. tenuis*) less acrylate after 6 days than juveniles at ambient temperature (Fig. 1c, d). Similar declines in acrylate in marine algae subjected to stressful conditions¹⁴ have been related to its antioxidant properties; acrylate and DMS are both extremely efficient scavengers of hydroxyl radicals and other reactive oxygen species (ROS)¹⁴. During thermal stress, the production of ROS by coral mitochondria (and *Symbiodinium*, when present) increases, damaging coral cells¹⁵. The observed decrease in acrylate concentrations in thermally stressed coral juveniles probably reflects its involvement in ROS detoxification, implicating DMSP and its breakdown products as functionally important in coral stress responses.

Further experimental studies involving adult corals, the building blocks of coral reefs and the most ecologically relevant life history stage in terms of DMSP production, confirm the importance of this

¹AIMS@JCU, and School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia. ²Australian Institute of Marine Science, PMB3, Townsville MC, Townsville, Queensland 4810, Australia. ³ARC Centre of Excellence for Coral Reef Studies, School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia. ⁴Australian National University, Canberra, Australian Capital Territory 2601, Australia. ⁵AIMS@JCU, and School of Pharmacy and Molecular Sciences, James Cook University, Townsville, Queensland 4811, Australia. ⁶School of Biological Sciences and Biotechnology, Murdoch University, Perth, Western Australia 6050, Australia. ⁷Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia. ⁸Oceans Institute, the University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia. [†]Present addresses: Natural Science and Public Health, Zayed University, PO Box 144534, Abu Dhabi, United Arab Emirates (D.A.); Kewalo Marine Lab, University of Hawai'i at Mānoa, 41 Ahui Street, Honolulu, Hawai'i 96813, USA (F.O.S.).

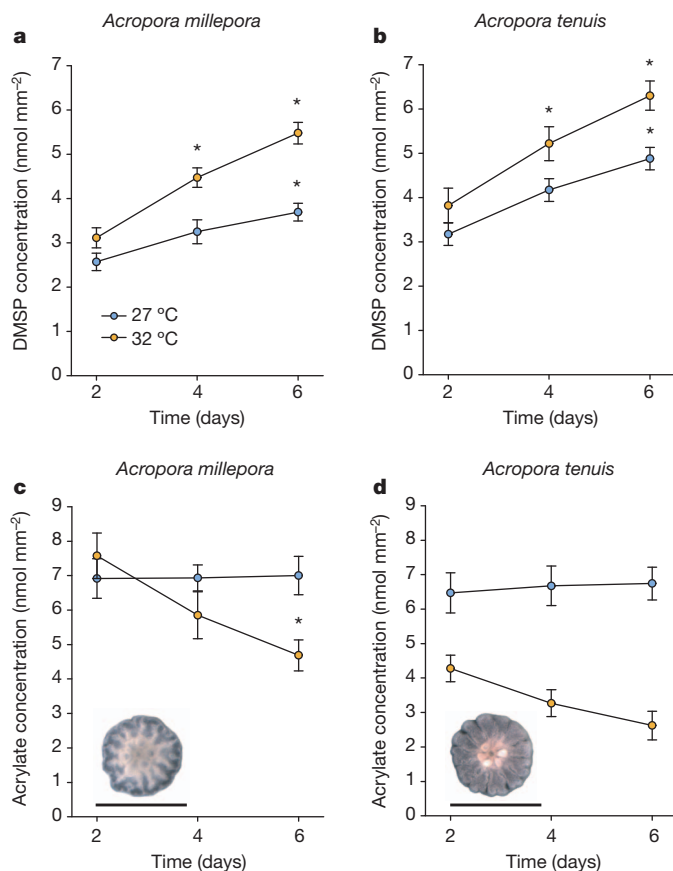


Figure 1 | Changes in DMSP and acrylate concentrations (mean \pm s.e.) in coral juveniles lacking photosynthetic symbionts ($n = 6$; 40 juveniles per replicate) through a 6-day period after coral settlement. a, b, Patterns are compared for two thermal regimes: ambient (27 °C, blue) and thermal stress (32 °C, orange), and for the coral species *Acropora millepora* (a) and *Acropora tenuis* (b). DMSP concentrations increase significantly through time in the juveniles exposed to ambient temperature (ANOVA simple main effect test, $F_{2,20} = 6.30$, $*P < 0.01$ for *A. millepora* and $F_{2,20} = 7.51$, $*P < 0.005$ for *A. tenuis*). Furthermore, DMSP concentrations in coral juveniles exposed to elevated temperature (32 °C) for 6 days were significantly higher than controls kept at ambient temperature (27 °C) (simple main effect test, $F_{1,10} = 27.68$, $*P < 0.0005$ for *A. millepora* and $F_{1,10} = 10.44$, $*P < 0.01$ for *A. tenuis*). c, d, Conversely, acrylate concentrations decreased significantly in juveniles of the corals *A. millepora* (c) and *A. tenuis* (d) when exposed to elevated temperature (simple main effect test, $F_{1,10} = 5.58$, $*P < 0.05$ for *A. millepora* and $F_{1,10} = 61.68$, $*P < 0.0005$ for *A. tenuis*). Scale bar, 1 mm.

compound in coral stress responses and the production of high concentrations of DMSP by the coral host. Colonies of *A. millepora* exposed to 32 °C for 10 days showed an 84% reduction in their *Symbiodinium* cell densities (Extended Data Fig. 1a and Fig. 2a–c), thus photosynthetic symbionts were unlikely to contribute significantly to DMSP production. Furthermore, transmission electron microscopy showed that 100% of the remaining *Symbiodinium* cells examined within coral tissues ($n = 272$ cells) were structurally compromised (Extended Data Fig. 2). In particular, thylakoid membranes within chloroplasts were completely disrupted and general cell structure was disordered (Extended Data Fig. 2), characteristics typical of advanced necrosis¹⁶. Consistent with these observations of dysfunctional cell structures, pulse amplitude modulation (PAM) fluorometry measurements confirmed significant reductions in their photochemical efficiency (ANOVA, $n = 12$, $P < 0.005$; Extended Data Fig. 1b and Extended Data Table 3). In further support of our conclusion that *Symbiodinium* were not contributing significantly to DMSP production, previous experiments on *Symbiodinium* cultures have shown that DMSP content per cell volume decreases under thermal stress¹⁷.

Despite the severe degradation and depletion of algal symbionts from coral tissues, thermally stressed corals contained 68% more DMSP (that is, 9 nmol mm⁻²) and concordantly 36% less acrylate than control colonies (Fig. 2). These results are in close accordance with those we found for coral juveniles. Taken together, our results provide conclusive evidence that the observed increases in DMSP concentrations in thermally stressed corals cannot be attributed to the activity of *Symbiodinium* cells. This implies that DMSP production is not restricted to juvenile life stages, but also occurs at high levels in adult reef-building corals subjected to thermal stress. In addition to its anti-oxidant role, DMSP is involved in several key cellular and ecological processes in reef-building corals, notably as an osmoprotectant during salinity fluctuations, but also as a signal molecule attracting specific bacteria which form microbial communities that are integrally associated with corals and underpin their health¹⁸. The multiple functions of this molecule might explain the ecological advantage of producing it directly.

To further support our conclusion that DMSP is produced by the coral host, we examined potential molecular mechanisms underlying their DMSP biosynthetic capabilities. In some marine algae, DMSP is produced from methionine via a pathway that involves the successive action of four different enzymes¹⁹, but until recently, very little was known about genes involved in its biosynthesis. A recent study identified candidate genes for each of the four steps of the biosynthesis pathway in the diatom *Fragilariopsis cylindrus*²⁰. We searched for potential orthologues of these genes in the comprehensive molecular resources available for two coral species, the *A. millepora* transcriptome²¹ and the *Acropora digitifera* genome²², and found that two genes identified in diatoms have clear orthologues in corals. These genes encode a NADPH-reductase and an AdoMet-dependant methyltransferase, which mediate the second and third steps of the biosynthesis process, respectively (Fig. 3a). The orthologous relationship between the corals' and diatom's genes is supported by best reciprocal blast hits and their mapping to the same OrthoMCL clusters²³ (Extended Data Table 4). We also identified two *Symbiodinium* sequences (from a comprehensive transcriptome assembly²⁴) belonging to the same OrthoMCL clusters as the reductase and methyltransferase identified in diatoms and corals (Extended Data Table 4). This indicates that the function of these enzymes in DMSP biosynthesis may be conserved between diatoms, alveolates, green plants and corals.

Results derived from the *A. millepora* transcriptome²¹ enabled us to estimate the level of expression of these two genes throughout five life stages, from embryos to adults. We found the gene encoding the reductase, a reversible step in the DMSP biosynthesis process⁵, to be highly expressed throughout all life-history stages (Fig. 3b). The phylogenetic distribution of its OrthoMCL cluster revealed 45 orthologues spread throughout all kingdoms (Extended Data Fig. 3), in line with the widespread presence of this enzyme among plants that do not produce DMSP⁵. In contrast, the methyltransferase mediates a non-reversible step regulating intracellular DMSP levels²⁵ and is believed to be specific to this pathway⁵. The gene encoding the methyltransferase was highly expressed in early coral life stages, but its expression decreased after settlement and remained relatively low in adult corals (Fig. 3c). This expression pattern correlates with the establishment of symbiosis with DMSP-producing *Symbiodinium* around the time of settlement. The OrthoMCL cluster corresponding to the methyltransferase has an unusually sparse phyletic pattern with only nine orthologues, including seven in photosynthetic organisms and two in other eukaryotes (Extended Data Fig. 3), indicating that corals are likely to be the exception, rather than the rule, in terms of DMSP production by marine invertebrates harbouring photosynthetic symbionts²⁶.

We show that the unparalleled levels of DMSP present in reef-building corals are not attributable solely to their endosymbiotic algae; instead, a significant fraction is produced by the coral animal. DMSP concentrations measured in coral juveniles devoid of symbiotic algae represented approximately half of those present in symbiont-bearing adult corals, indicating that the coral animal contributes extensively to

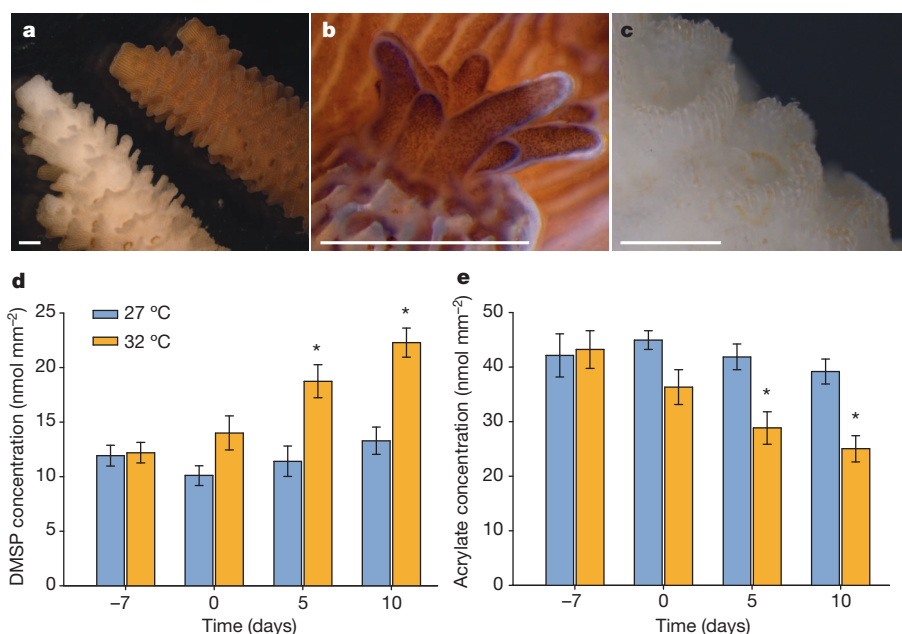


Figure 2 | Effects of thermal stress on adult colonies of the coral *Acropora millepora*. Temperature was raised gradually over 7 days to 32 °C, mimicking a realistic thermal stress scenario, therefore thermal stress commenced on day 0. The total duration of the experiment was 17 days. **a**, Micrograph of representative coral nubbins showing visual differences in the density of *Symbiodinium* cells present in the tissues of *A. millepora* maintained under control (27 °C, brown, right) or thermal stress (32 °C, white, left) conditions for 10 days. **b**, **c**, Higher magnification micrographs of coral polyps showing visual differences in tissue colour as intracellular *Symbiodinium* densities decline between 27 °C (**b**) and 32 °C (**c**) (scale bars, 1 mm). **d**, Changes in DMSP concentration (mean \pm s.e.) in adult corals ($n = 12$) exposed to control (27 °C) and thermal stress (32 °C) treatments (ANOVA simple main effect test, $F_{1,22} = 10.79$, $*P < 0.005$). **e**, Corresponding changes in acrylate concentration (mean \pm s.e.) in adult corals ($n = 12$) exposed to control (27 °C) and thermal stress (32 °C) treatments (ANOVA simple main effect test, $F_{1,22} = 8.4$, $*P < 0.01$).

the DMSP pool produced on coral reefs. Surveys over Australia identified the Great Barrier Reef as a significant hotspot for the emission of sulphur aerosol particles^{27,28}. The Great Barrier Reef is the largest biological structure on the planet and the release of these particles along its 2,600-km length could constitute a major source of cloud condensation nuclei. Coral-reef-derived sulphur aerosol emission might therefore have a central role in cloud formation in areas of the world with high coral densities, such as the Great Barrier Reef and the Coral

Triangle^{11,29}. Considering declining trends in coral cover³⁰ and predicted increases in coral mortality worldwide caused by anthropogenic stressors, the associated decline in sulphur aerosol production from coral reefs may further destabilize local climate regulation and accelerate degradation of this globally important and diverse ecosystem.

METHODS SUMMARY

Coral juvenile sampling. Coral larvae were raised for 12 days in 0.5- μ m-filtered seawater (FSW), subsequently washed three times in 0.2- μ m FSW and settled in sterile 6-well plates (8 plates per species, 40 larvae per well; each well filled with 10 ml of 0.2- μ m FSW). Eight hours after settlement, the plates were separated between two treatments: 4 plates per species were incubated at 27 °C (control temperature), and the other 4 plates were ramped to 32 °C over 5 h (thermal stress treatment). All plates were kept in the dark (to prevent the growth of potential photosynthetic organisms) and 6 random wells were sampled every 2 days.

DNA extractions and PCR amplification. At each time point, the contents of two wells were scraped into a 2-ml tube using a scalpel blade, snap frozen with liquid nitrogen, and used for total DNA extraction. Multiple sets of primers were subsequently used to target the *Symbiodinium*, coral and other photosynthetic organisms (Extended Data Tables 1 and 5).

Identification of candidate genes. Orthology between coral, *Symbiodinium* and diatom genes was inferred based on best reciprocal BLAST hits. Orthologues in other species were identified using release 5 (March 2011) of the OrthoMCL database³. Database for *A. digitifera* genome: http://marinegenomics.oist.jp/genomes/download?project_id=3 annotation version 1.0. Database for *A. millepora* transcriptome: NCBI Transcriptome Shotgun Assembly (TSA)²¹.

Adult thermal stress experiment. Ten coral colonies were split to give a total of 24 fragments, each comprising approximately 25 branches (nubbins). Fragments were arranged in eight indoor tanks in a randomized block design. After 2 weeks of acclimatization, the temperature in four tanks was ramped-up slowly to 32 °C, whereas the other four tanks were maintained at 27 °C. Each temperature treatment contained a total of 12 fragments.

qNMR analysis. Adult and juvenile corals were extracted and analysed in accordance with the qNMR technique described previously¹³.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 July; accepted 18 September 2013.

Published online 23 October 2013.

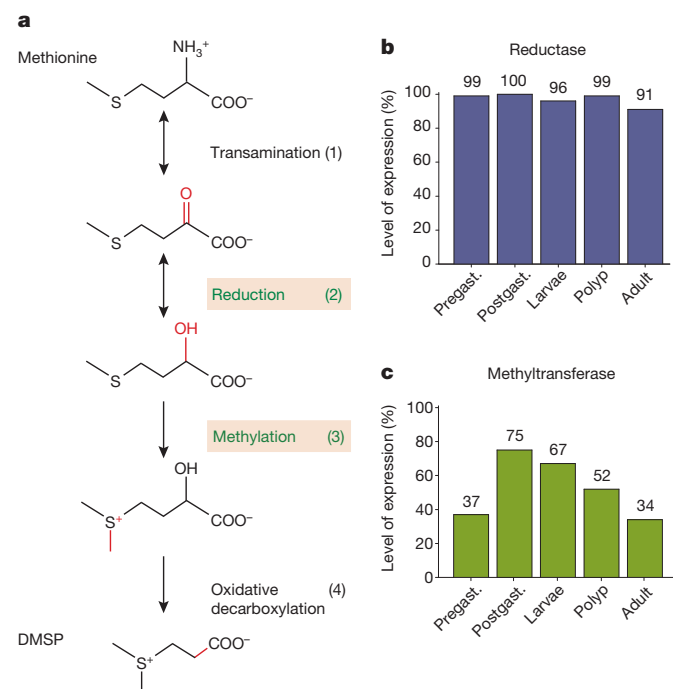


Figure 3 | Putative DMSP biosynthesis pathway and gene expression of coral orthologues. **a**, Pathway of DMSP biosynthesis present in marine algae¹⁹ (the first two steps are reversible). The proposed diatom genes encoding the second and third steps (in green) have orthologues in *Acropora millepora* and *Acropora digitifera* genomes (see Extended Data Table 4). **b**, **c**, Gene expression percentiles of the two putative genes involved in DMSP biosynthesis in *A. millepora* through five coral life stages, from embryos (pregast., pregastrula; postgast., postgastrula) to adults, based on transcriptomic data from ref. 21.

- Broadbent, A. D. & Jones, G. B. DMS and DMSP in mucus ropes, coral mucus, surface films and sediment pore waters from coral reefs in the Great Barrier Reef. *Mar. Freshw. Res.* **55**, 849–855 (2004).
- Broadbent, A. D., Jones, G. B. & Jones, R. J. DMSP in corals and benthic algae from the Great Barrier Reef. *Estuar. Coast. Shelf Sci.* **55**, 547–555 (2002).

3. Ayers, G. P. & Gras, J. L. Seasonal relationship between cloud condensation nuclei and aerosol methanesulphonate in marine air. *Nature* **353**, 834–835 (1991).
4. Vallina, S. M. & Simo, R. Strong relationship between DMS and the solar radiation dose over the global surface ocean. *Science* **315**, 506–508 (2007).
5. Stefels, J. Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *J. Sea Res.* **43**, 183–197 (2000).
6. Todd, J. D. *et al.* Structural and regulatory genes required to make the gas dimethylsulfide in bacteria. *Science* **315**, 666–669 (2007).
7. Sievert, S. M., Kiene, R. P. & Schulz-Vogt, H. N. The sulfur cycle. *Oceanography* **20**, 117–123 (2007).
8. Quinn, P. K. & Bates, T. S. The case against climate regulation via oceanic phytoplankton sulphur emissions. *Nature* **480**, 51–56 (2011).
9. Yost, D. M., Jones, R., Rowe, C. L. & Mitchelmore, C. L. Quantification of total and particulate dimethylsulfoniopropionate (DMSP) in five Bermudian coral species across a depth gradient. *Coral Reefs* **31**, 561–570 (2012).
10. Yost, D. M. & Mitchelmore, C. L. Determination of total and particulate dimethylsulfoniopropionate (DMSP) concentrations in four scleractinian coral species: a comparison of methods. *J. Exp. Mar. Biol. Ecol.* **395**, 72–79 (2010).
11. Fischer, E. & Jones, G. B. Atmospheric dimethylsulphide production from corals in the Great Barrier Reef and links to solar radiation, climate and coral bleaching. *Biogeochemistry* **110**, 31–46 (2012).
12. Veron, J. E. N. *Corals of the World* Vol. 1–3 (Australian Institute of Marine Science, 2000).
13. Tapiolas, D. M., Raina, J. B., Lutz, A., Willis, B. L. & Motti, C. A. Direct measurement of dimethylsulfoniopropionate (DMSP) in reef-building corals using quantitative nuclear magnetic resonance (qNMR) spectroscopy. *J. Exp. Mar. Biol. Ecol.* **443**, 85–89 (2013).
14. Sunda, W., Keiber, D. J., Kiene, R. P. & Hunstman, S. An antioxidant function for DMSP and DMS in marine algae. *Nature* **418**, 317–320 (2002).
15. Lesser, M. P. in *Coral Reefs: An Ecosystem in Transition* (eds Dubinsky, Z. & Stambler, N.) 405–419 (Springer, 2011).
16. Strychar, K. B., Sammarco, P. W. & Piva, T. J. Apoptotic and necrotic stages of *Symbiodinium* (Dinophyceae) cell death activity: bleaching of soft and scleractinian corals. *Phycologia* **43**, 768–777 (2004).
17. McLendon, A. L. & DiTullio, G. R. Effect of increased temperature on dimethylsulfoniopropionate (DMSP) concentration and methionine synthase activity in *Symbiodinium microadriaticum*. *Biogeochemistry* **110**, 17–29 (2012).
18. Raina, J. B., Dinsdale, E. A., Willis, B. L. & Bourne, D. G. Do the organic sulfur compounds DMSP and DMS drive coral microbial associations? *Trends Microbiol.* **18**, 101–108 (2010).
19. Gage, D. A. *et al.* A new route for synthesis of dimethylsulfoniopropionate in marine algae. *Nature* **387**, 891–894 (1997).
20. Lyon, B. R., Lee, P. A., Bennett, J. M., DiTullio, G. R. & Janech, M. G. Proteomic analysis of a sea-ice diatom: salinity acclimation provides new insight into the dimethylsulfoniopropionate production pathway. *Plant Physiol.* **157**, 1926–1941 (2011).
21. Moya, A. *et al.* Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO₂-driven acidification during the initiation of calcification. *Mol. Ecol.* **21**, 2440–2454 (2012).
22. Shinzato, C. *et al.* Using the *Acropora digitifera* genome to understand coral response to environmental change. *Nature* **476**, 320–323 (2011).
23. Chen, F., Mackey, A. J., Vermunt, J. K. & Roos, D. S. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS ONE* **2**, e383 (2007).
24. Bayer, T. *et al.* *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS ONE* **7**, e35269 (2012).
25. Ito, T., Asano, Y., Tanaka, Y. & Takabe, T. Regulation of biosynthesis of dimethylsulfoniopropionate and its uptake in sterile mutant of *Ulva pertusa* (Chlorophyta). *J. Phycol.* **47**, 517–523 (2011).
26. Van Alstyne, K. L., Dominique, V. J. & Muller-Parker, G. Is dimethylsulfoniopropionate (DMSP) produced by the symbionts or the host in an anemone-zooxanthellae symbiosis? *Coral Reefs* **28**, 167–176 (2009).
27. Bigg, E. K. & Turvey, D. E. Sources of atmospheric particles over Australia. *Atmos. Environ.* **12**, 1643–1655 (1978).
28. Modini, R. L. *et al.* New particle formation and growth at a remote, sub-tropical coastal location. *Atmos. Chem. Phys.* **9**, 7607–7621 (2009).
29. Leahy, S. M., Kingsford, M. J. & Steinberg, C. R. Do clouds save the Great Barrier Reef? Satellite imagery elucidates the cloud-SST relationship at the local scale. *PLoS ONE* **8**, e70400 (2013).
30. De'ath, G., Fabricius, K. E., Sweatman, H. & Puotinen, M. The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc. Natl Acad. Sci. USA* **109**, 17995–17999 (2012).

Acknowledgements The authors would like to thank D. Yellowlees, R. Stocker, M. Garren, A. Johnston, W. Dunlap, H. Harrison, P. Warner and E. Botté for valuable comments on the manuscript. We specially thank K. Ritchie for her advice and encouragement. We also thank J. Hicks, P. Barron (Bruker Biospin), A. Negri, T. Harder, J. Tebben, M. Logan and J. Pollock for their assistance. This work was supported by the AMMRF Centre for Microscopy, Characterisation and Analysis (UWA), the ARC Centre of Excellence for Coral Reef Studies and AIMS.

Author Contributions J.-B.R. and A.L. designed the experiments. J.-B.R., C.A.M., D.A. and D.M.T. performed the juvenile experiment. J.-B.R., C.A.M., A.L., F.O.S. and D.M.T. performed the adult experiment. J.-B.R., P.L.C., C.A.M. and D.M.T. analysed the results. S.F. and J.-B.R. identified the candidate genes. J.-B.R. and B.L.W. wrote the manuscript. All authors edited the manuscript before submission.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-B.R. (j.raina@aims.gov.au).

METHODS

Adult corals

Thermal stress experiment. *Acropora millepora* colonies ($n = 10$) were collected from Pelorus Island, Great Barrier Reef, Australia ($18^{\circ}33' \text{ S}/146^{\circ}29' \text{ E}$) and transferred to the Australian Institute of Marine Science. Coral colonies were fragmented to give a total of 24 fragments, each comprising approximately 25 branches (nubbins). Fragments were arranged in eight indoor tanks in a randomized block design, resulting in the allocation of 12 coral fragments to each of the control and thermal stress temperature treatments (27°C and 32°C , respectively). All tanks were continuously supplied with fresh 1- μm filtered seawater (FSW), which was maintained at 27°C ($\pm 0.1^{\circ}\text{C}$) via computer control using a flow-through system at a rate of 1.5 l min^{-1} . Ultraviolet-filtered lights were mounted above each tank and provided an average underwater light intensity of $350 \mu\text{E}$ over a 12/12 h light/dark cycle (400 W metal halide lamps, BLV), typical of light intensities recorded at the collection site. The fragments were acclimatized for 2 weeks before starting the experiment. Seawater temperatures in four tanks were slowly and continuously ramped to 32°C ($\pm 0.05^{\circ}\text{C}$) over a 7-day period, via computer control, while the remaining four control tanks were maintained at 27°C for the entire duration of the experiment.

Coral nubbins were sampled four times during the experiment: before any temperature changes when both treatments were at 27°C ($t = -7$); once the 32°C target temperature had been reached in the thermal stress treatment ($t = 0$); after 5 days at 32°C , when the first physiological effects of temperature stress were visible ($t = 5$); and after 10 days at 32°C when all colonies in the 32°C treatment were completely bleached ($t = 10$). At each time point, one coral nubbin (approximately 50 mm in length) was collected from each coral fragment ($n = 24$) and immediately transferred to a tube containing 2 ml of HPLC-grade methanol for quantitative nuclear magnetic resonance (qNMR) analysis. Another coral nubbin was collected from each coral fragment ($n = 24$) to evaluate *Symbiodinium* densities. In addition, one coral nubbin was collected at each time point from four different coral fragments, transferred directly into fixative (1.25% glutaraldehyde + 0.5% paraformaldehyde in 0.2- μm FSW) and stored at 4°C until processed for structural investigations by transmission electron microscopy (TEM).

Quantitative NMR analysis. The coral nubbins were extracted in methanol for 2 h with sonication followed by a second extraction with an additional 1 ml of HPLC-grade methanol for 10 min. The two extracts were pooled and dried using a vacuum-centrifuge then re-suspended in a mixture of deuterium oxide (D_2O , D 99.8%, 250 μl) and deuterated methanol (CD_3OD , D 99.8%, 750 μl) (Cambridge Isotope Laboratories). A 700 μl aliquot of the particulate-free extract was transferred into a 5-mm Norell 509-UP-7 NMR tube (Norell Inc.) and analysed immediately by ^1H NMR.

^1H NMR spectra were recorded on a Bruker Avance 600 MHz NMR spectrometer with TXI 5-mm probe and quantification performed using the ERETIC method¹³. This technique generates an internal electronic reference signal, calibrated using commercial stock solutions of 4 mM acrylate and DMSP. The concentrations of DMSP and acrylate were determined by integration of their respective signals in a 0.10 p.p.m. window¹³.

***Symbiodinium* densities.** Freshly collected coral nubbins were airbrushed (80 lb in⁻²) in individual plastic bags in 4 ml of 0.2- μm FSW. The slurry was homogenized to breakdown aggregates and centrifuged at 3,000 r.c.f. (relative centrifugal force). The supernatant was removed and the pellet re-suspended in 1 ml of 10% formalin. Homogeneous extracts were placed on a hemocytometer (depth 0.1 mm) and *Symbiodinium* cells were counted under a light microscope (eight technical replicates were averaged per sample).

***Symbiodinium* genotype.** *In hospite Symbiodinium* populations of each experimental coral fragment were characterized on the basis of sequence differences in the nuclear ribosomal DNA internal transcribed spacer 1 region using single-strand conformation polymorphism (SSCP) analysis³¹. Total DNA was extracted using a modified protocol³². *Symbiodinium* ITS1 region was amplified with fluorescently labelled Sym ITS1 PCR primers. Genotype was determined using SSCP with known reference samples running alongside experimental samples and scored manually using gel images. All colonies contained only *Symbiodinium* type C2 (GenBank accession AF380552) SSCP profiles from all samples were single bands identical to the reference).

Surface-area calculation. Coral skeletons remaining after samples were extracted for qNMR analyses and *Symbiodinium* densities were lyophilized overnight and their surface area determined using a wax dipping technique³³. The surface area of each individual nubbin was used to normalize the qNMR and *Symbiodinium* data.

Transmission electron microscopy. Fixed coral nubbins were decalcified in a formic acid:fixative mixture (1:3), with the solution changed every 12 h until complete dissolution of the skeleton. Three individual polyps per sample were post-fixed in osmium and subsequently dehydrated with increasing concentrations of ethanol followed by dry acetone. Dehydrated samples were infiltrated in increasing

concentrations of Araldite resin before being cured for 24 h at 60°C . Longitudinal sections 90-nm thick were collected on copper grids and imaged at 120 kV in a JEOL 2100 TEM.

Pulse amplitude modulation (PAM) fluorometry measurements. Photosystem II (PSII) photochemical efficiency was measured with a Diving-PAM (Walz) on three random nubbins per coral fragment. Minimum and maximum fluorescence (F_0 and F_M) were recorded daily, 2 h before the start of the light cycle. PS II photochemical efficiency was expressed as maximum quantum yields $((F_M - F_0)/F_M) = (F_V/F_M)$.

Data analyses. All data were square root transformed and no significant tank effect was detected for the DMSP, acrylate and PAM fluorometry data (nested ANOVA, $P > 0.05$) (Statistica 7, Statsoft). The sample size was chosen based on results from a pilot study. Repeated measures ANOVA were carried out on the time series data (data met all assumptions of the test). Simple main effect tests³⁴ were used to compare the results between the two temperature treatments, and between temperatures at each time point (Extended Data Table 3). This statistical technique was used to minimize the number of multiple comparisons (focusing only on comparisons of interest), decreasing the likelihood of type I error.

Coral juveniles

Sample collections. Colonies of *A. millepora* ($n = 10$) and *A. tenuis* ($n = 3$) were collected from Orpheus Island, Great Barrier Reef, Australia ($18^{\circ}34' \text{ S}/146^{\circ}30' \text{ E}$) and transferred to the Australian Institute of Marine Science outdoor aquarium facility 4 days before the predicted spawning event in November 2011. One hour before spawning, the colonies were isolated in 70 l tanks with 1- μm FSW. Gametes were collected from the surface of these tanks and fertilized in separate 70-l tanks with FSW. After fertilization, embryos were gently rinsed three times by transferring to new containers and were subsequently transferred to 500-l tanks (containing 0.5- μm FSW) where they were kept through larval development. After 12 days, *Symbiodinium*-free coral larvae were collected using a 1 μm mesh net and washed three times in 0.2- μm FSW. Larvae were subsequently settled in sterile 6-well plates (8 plates per species, 40 larvae per well; each well filled with 10 ml of 0.2- μm FSW). Eight hours after settlement, the plates were separated between two temperature regimes: 4 plates per species were incubated at 27°C (control temperature), and the other 4 plates were ramped to 32°C over 5 h (thermal stress treatment); all plates were maintained in the dark (to prevent the growth of potential photosynthetic organisms). Settled juveniles were incubated at their respective temperature treatments and six random wells were sampled every 2 days. The size of the sampled juveniles was measured using a motorized stereomicroscope (Leica MZ16A, Leica Microsystems); sizes did not vary significantly between day 2 and day 6 with an average size of 0.79 mm^2 (± 0.05) for *A. millepora* juveniles and of 0.82 mm^2 (± 0.07) for *A. tenuis* juveniles.

qNMR analysis. After the required incubation time, seawater was removed from each well with a pipette, and a sterile cotton bud was used to soak up any residual seawater, taking care not to disturb the settled juvenile corals. Juveniles in six wells were extracted by adding 300 μl of deuterated methanol (CD_3OD) to each well, followed by 30 s of gentle shaking; 200 μl of this extract was transferred into a 3-mm Bruker MATCH NMR tube and analysed immediately. In addition, negative control wells without settled juveniles were extracted following the same procedure. The concentrations of DMSP and acrylate were normalized initially to the number of settled coral juveniles in the respective well. They were then normalized to the averaged surface area of the juveniles (note that juveniles were approximated to perfect circles).

DNA extractions and PCR amplification. At each time point, the contents of two wells were scraped into a 2-ml tube using a scalpel blade, snap frozen with liquid nitrogen, and used for total DNA extraction, according to methods in ref. 35. Multiple sets of primers were subsequently used to target different taxonomic groups: *Symbiodinium*³¹, coral³⁶ and general primers for the algal 23S rDNA plastid³⁷ and chloroplast DNA³⁸ (Extended Data Table 1). Furthermore, three other DNA markers were used to target microbes: Bacteria³⁹, Archaea⁴⁰ and Fungi⁴¹ (Extended Data Table 1).

The PCR consisted of 1 μl of DNA template (dilution series from 1 to 10^{-5} of the original concentrations), 10 μl of buffer containing dNTP and MgCl_2 (Bioline), 1.5 μl of each primer (10 μM), and 0.5 μl of *Taq* polymerase (Bioline), adjusted to a final volume of 50 μl with sterile MilliQ water. Amplified PCR products were visualized by electrophoresis on 1% agarose gel stained with ethidium bromide.

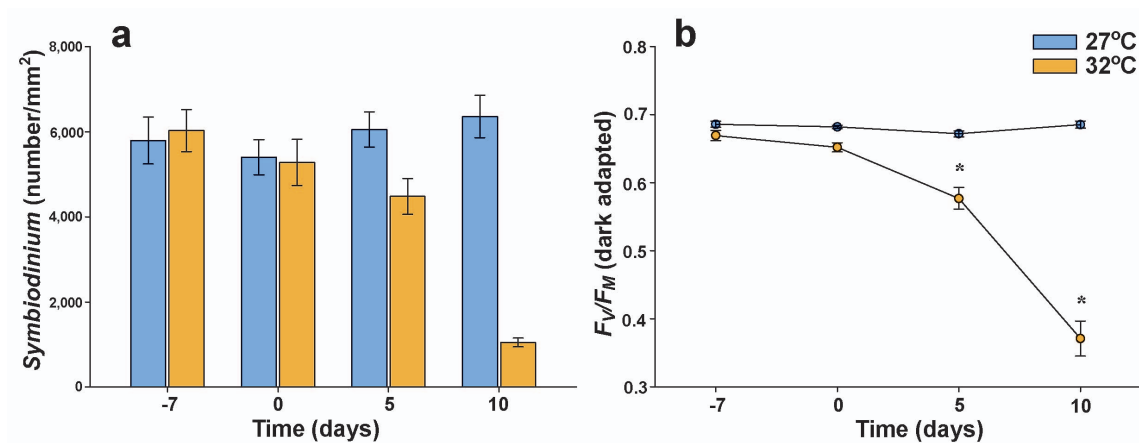
Clone libraries construction. PCR products from bacterial 16S ribosomal RNA gene were purified using a QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. Purified DNA was ligated into a TOPO-TA cloning vector (Invitrogen) and competent cells were transformed following the manufacturer's instructions. Four samples, derived from the last experimental time point (6 days), were analysed (*A. millepora* at 27°C ; *A. millepora* at 32°C ; *A. tenuis* at 27°C ; and *A. tenuis* at 32°C). A total of 48 clones were sequenced per sample (Macrogen Inc.) and chimaeric sequences were removed from subsequent analysis.

The nucleotide sequences obtained have been deposited in GenBank database under the accession numbers (KF619251 to KF619442).

Identification of candidate genes. Orthology between coral and diatom genes was inferred on the basis of best reciprocal BLAST hits⁴². Orthologues in other species were identified using release 5 (March 2011) of the OrthoMCL database²³. Database for *A. digitifera* genome: http://marinegenomics.oist.jp/genomes/download?project_id=3 annotation version 1.0; for *A. millepora* transcriptome: NCBI Transcriptome Shotgun Assembly (TSA)²¹.

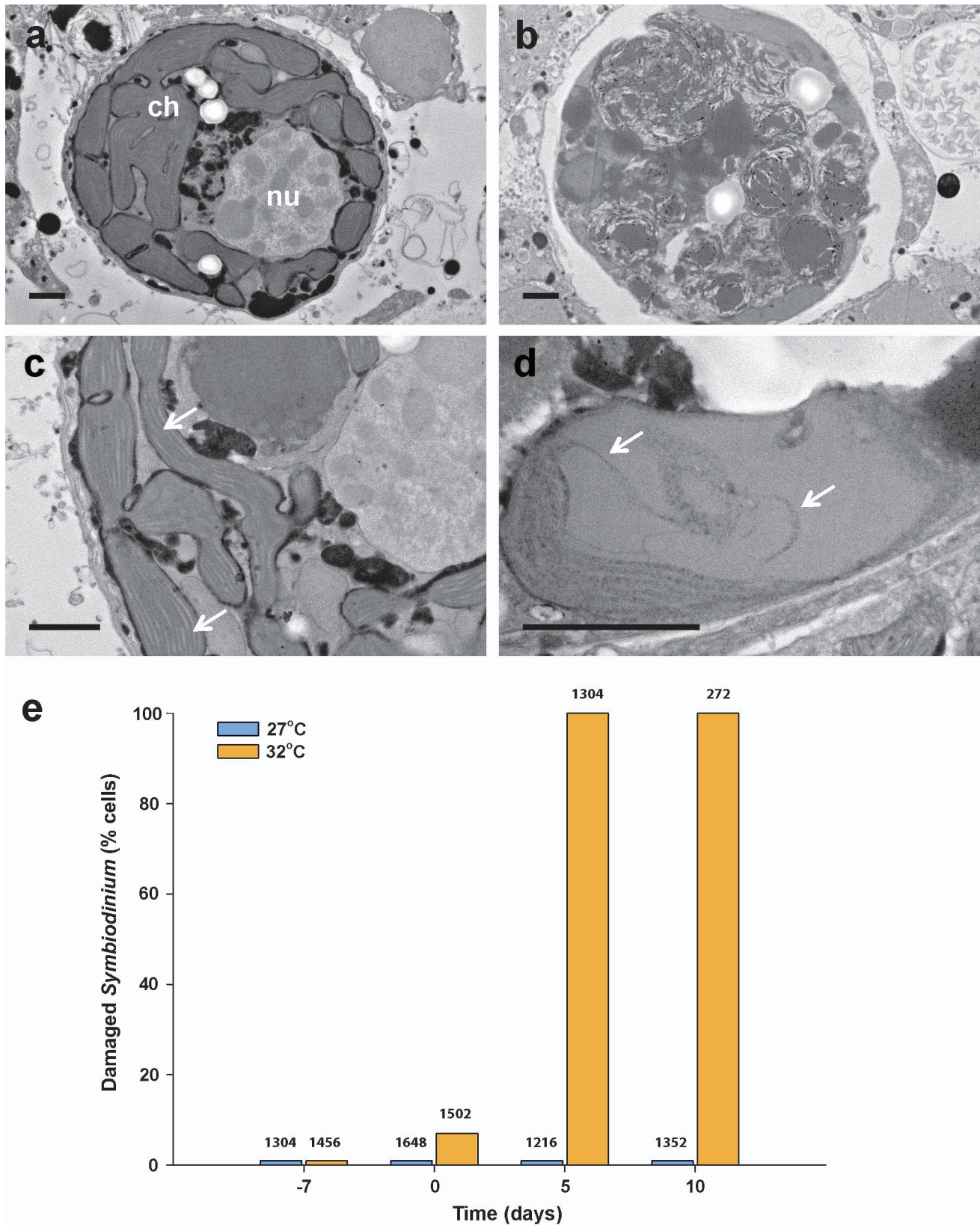
Data analyses. Repeated measure ANOVA were performed on the normalized DMSP and acrylate concentrations (data met all assumptions of the test, except the acrylate data from *A. tenuis* that violated sphericity assumption; in that unique case, Greenhouse–Geisser correction was applied to the degrees of freedom). Simple main effect tests³⁴ were then used to compare the results from the temperature treatment within species (Extended Data Table 2).

31. van Oppen, M. J. H., Palstra, F. P., Piquet, A. M. T. & Miller, D. J. Patterns of coral-dinoflagellate associations in *Acropora*: significance of local availability and physiology of *Symbiodinium* strains and host-symbiont selectivity. *Proc. R. Soc. Lond. B* **268**, 1759–1767 (2001).
32. Wilson, K. *et al.* Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism. *Aquaculture* **204**, 297–309 (2002).
33. Veal, C. J., Holmes, G., Nunez, M., Hoegh-Guldberg, O. & Osborn, J. A comparative study of methods for surface area and three dimensional shape measurement of coral skeletons. *Limnol. Oceanogr. Methods* **8**, 241–253 (2010).
34. Quinn, G. P. & Keough, M. J. *Experimental Design and Data Analysis for Biologists* (Cambridge Univ. Press, 2002).
35. Bourne, D. & Munn, C. Diversity of bacteria associated with the coral *Pocilloporada micromis* from the Great Barrier Reef. *Environ. Microbiol.* **7**, 1162–1174 (2005).
36. Suzuki, G., Hayashibara, T., Shirayama, Y. & Fukami, H. Evidence of species-specific habitat selectivity of *Acropora* corals based on identification of new recruits by two molecular markers. *Mar. Ecol. Prog. Ser.* **355**, 149–159 (2008).
37. Sherwood, A. R., Chan, Y. L. & Presting, G. G. Application of universally amplifying plastid primers to environmental sampling of a stream periphyton community. *Mol. Ecol. Res.* **8**, 1011–1014 (2008).
38. Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **17**, 1105–1109 (1991).
39. Marchesi, J. R. *et al.* Design and evaluation of useful bacterium-specific primers that amplify genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **64**, 795–799 (1998).
40. Takai, K. & Horikoshi, K. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl. Environ. Microbiol.* **66**, 5066–5072 (2000).
41. White, T. J., Bruns, T., Lee, S. & Taylor, J. *Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. PCR Protocols: a Guide to Methods and Applications* (ed. Innis, M.) (Academic, 1990).
42. Moreno-Hagelsieb, G. & Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**, 319–324 (2008).
43. Raina, J. B., Tapiolas, D., Willis, B. L. & Bourne, D. G. Coral-associated bacteria and their role in the biogeochemical cycling of sulfur. *Appl. Environ. Microbiol.* **75**, 3492–3501 (2009).
44. González, J. M. *et al.* Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl. Environ. Microbiol.* **66**, 4237–4246 (2000).
45. Todd, J. D. *et al.* DddQ, a novel, cupin-containing, dimethylsulfoniopropionatelyase in marine roseobacters and in uncultured marine bacteria. *Environ. Microbiol.* **13**, 427–438 (2010).
46. Todd, J. D. *et al.* Structural and regulatory genes required to make the gas dimethylsulfide in bacteria. *Science* **315**, 666–669 (2007).
47. Gómez-Consarnau, L., Lindh, M. V., Gasol, J. M. & Pinhassi, J. Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environ. Microbiol.* **14**, 2361–2378 (2012).
48. Curson, A. R. J., Rogers, R., Todd, J. D., Brearley, C. A. & Johnston, A. W. B. Molecular genetic analysis of a dimethylsulfoniopropionatelyase that liberates the climate-changing gas dimethylsulfide in several marine α -proteobacteria and *Rhodobacter sphaeroides*. *Environ. Microbiol.* **10**, 757–767 (2008).
49. Todd, J. D., Curson, A. R. J., Nicholson, P. & Johnston, A. W. B. The *dddP* gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. *Environ. Microbiol.* **11**, 1376–1385 (2009).
50. Howard, E. C. *et al.* Bacterial taxa that limit sulfur flux from the ocean. *Science* **314**, 649–652 (2006).
51. Schäfer, H. Isolation of *Methylophaga* spp. from marine dimethylsulfide degrading enrichment cultures and identification of polypeptides induced during growth on dimethylsulfide. *Appl. Environ. Microbiol.* **73**, 2580–2591 (2007).
52. Vogt, C. & Fischer, U. Influence of reduced inorganic sulfur compounds and oxygen on DMS oxidation and DMS reduction by the marine purple “nonsulfur” bacterium *Rhodovulum sulfidophilum* strain W4. *Microbiol. Res.* **153**, 219–226 (1998).



Extended Data Figure 1 | Density and photosynthetic efficiency (mean \pm s.e.) of *Symbiodinium* cells within adult colonies of the coral *Acropora millepora* maintained under control (27 °C) or thermal stress (32 °C) conditions for 10 days. a, Density of *Symbiodinium* cells in the same

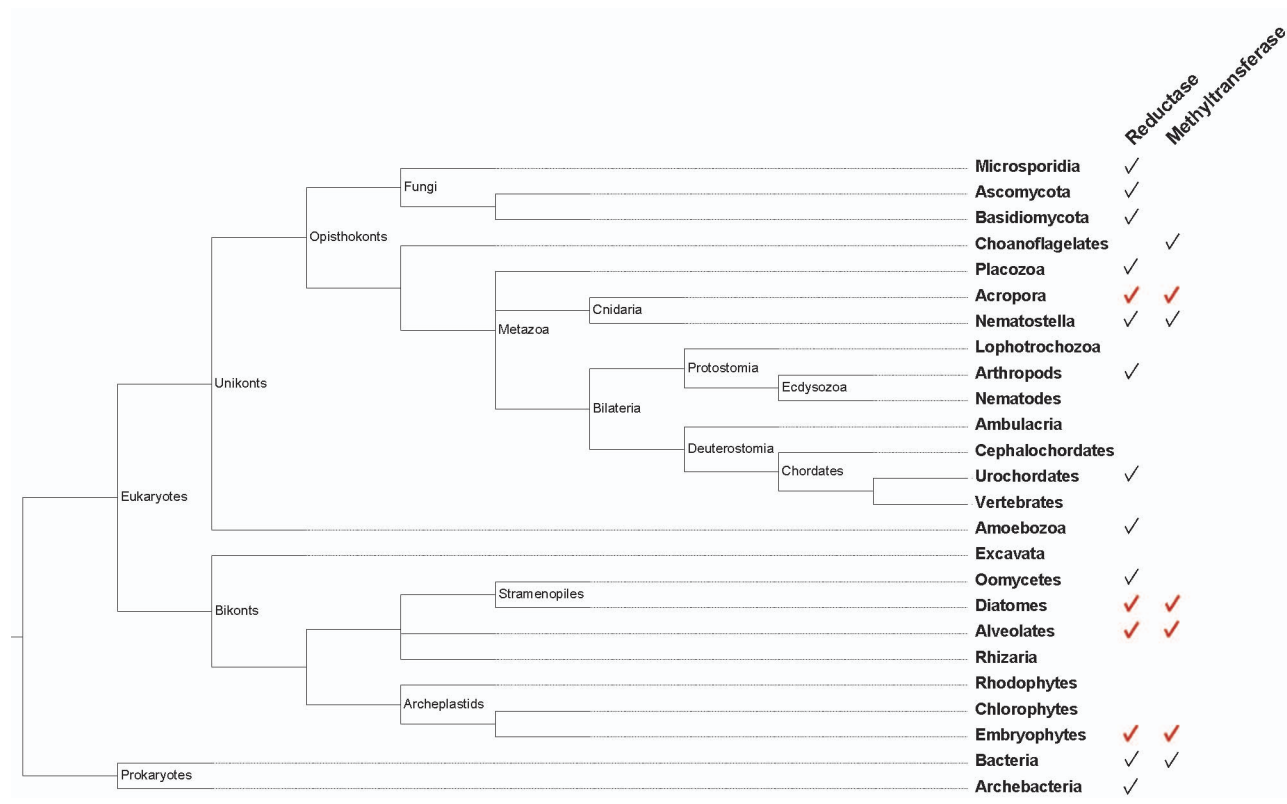
coral fragments through time. b, Comparison of photosystem II photochemical efficiency (maximum quantum yields: F_v/F_m) through time (repeated measure ANOVA, * $P < 0.001$; post-hoc simple main effect test, * $P < 0.01$). See also Extended Data Table 3.



Extended Data Figure 2 | Representative transmission electron micrographs showing the effects of thermal stress on the internal structure of endosymbiotic *Symbiodinium* cells associated with the coral *Acropora millepora*. a, c, *Symbiodinium* cells after 10 days at 27°C, showing intact cell structures (a) and intact thylakoid membranes of chloroplasts (arrows), the photosynthetic centre of cells (c). b, d, *Symbiodinium* cells after 10 days at

32°C, showing structurally degraded cells (b) with highly disrupted thylakoid membranes (arrows) (d). Scale bars, 1 µm. ch, chloroplast; nu, nucleus.

e, Percentage of structurally damaged *Symbiodinium* cells within adult tissue throughout the thermal stress experiment. The numbers above the bars refer to the total number of *Symbiodinium* cells observed.



Extended Data Figure 3 | Phylogenetic distribution of the reductase and methyltransferase orthologues (OrthoMCL groups OG5_131390 and OG5_156314, respectively). Note the unusually sparse distribution of OG5_156314. In red: co-occurrence of these two enzymes occurs

predominantly in DMSP-producing organisms. The only species of bacteria in the OrthoMCL database where these two enzymes occur simultaneously is the marine cyanobacterium *Synechococcus*.

Extended Data Table 1 | Primer pairs used to target the genomic DNA of various microorganisms possibly responsible for DMSP production to assess presence in coral juveniles

Target organisms	Target region	Primer name	Amplicon size	Annealing temperature	Number of cycles	Reference	Amplification
Coral	mtDNA	RNS2/GR	700 bp	54°C	30	³⁶	Yes
<i>Symbiodinium</i>	ITS1	ITSFP/ITSRP	350 bp	59°C	30	³¹	No
Algae	23S rDNA plastid	p23SrVf1/Vr1	410 bp	Touchdown PCR (66 to 58°C)	35	³⁷	No
Photosynthetic organisms	cpDNA	a/b	Variable	55°C	35	³⁸	No
Photosynthetic organisms	cpDNA	c/d	Variable	55°C	35	³⁸	No
Photosynthetic organisms	cpDNA	e/f	Variable	55°C	35	³⁸	No
Bacteria	16S rRNA	63f/1387r	1300 bp	55°C	30	³⁹	Yes
Fungi	ITS1	ITS1f/ITS4	Variable	55°C	30	⁴¹	No
Archaea	16 rRNA	349f/806r	457 bp	59°C	30	⁴⁰	No

mtDNA, mitochondrial DNA; cpDNA, chloroplast DNA; ITS, internal transcribed spacer; bp, base pairs.

Extended Data Table 2 | Sums of squares (SS), mean squares (MS) and significance levels for ANOVAs of the *Symbiodinium*-free juvenile experiment

a DMSP in <i>Acropora millepora</i>					
Effect	SS	df	MS	F	<i>p</i>
Intercept	132.6612	1	132.6612	6959.515	0.000000
Temperature	0.8075	1	0.8075	42.3610	0.000068
Time 2	0.0788	1	0.0788	4.1343	0.069420
Time 4	0.3015	1	0.3015	15.8185	0.002612
Time 6	0.5277	1	0.5277	27.6862	0.000367
Error	0.1906	10	0.0191		
Time	1.2373	2	0.6186	24.7230	0.000004
Time × Temperature	0.1007	2	0.0504	2.0120	0.159840
27°C	0.3160	2	0.1580	6.3099	0.007500
32°C	1.0220	2	0.5110	20.4073	0.000015
Error	0.5005	20	0.0250		
DMSP in <i>Acropora tenuis</i>					
Effect	SS	df	MS	F	<i>p</i>
Intercept	162.4001	1	162.4001	6250.471	0.000000
Temperature	0.5032	1	0.5032	19.368	0.001333
Time 2	0.0858	1	0.0858	3.3033	0.099181
Time 4	0.1720	1	0.1720	6.6189	0.027760
Time 6	0.2715	1	0.2715	10.4499	0.008981
Error	0.2598	10	0.0260		
Time	1.5042	2	0.7521	19.8690	0.000018
Time × Temperature	0.0261	2	0.0130	0.3440	0.712926
27°C	0.5677	2	0.2839	7.5112	0.003688
32°C	0.9625	2	0.4812	12.7343	0.000271
Error	0.7571	20	0.0378		
b Acrylate in <i>Acropora millepora</i>					
Effect	SS	df	MS	F	<i>p</i>
Intercept	230.3004	1	230.3004	1830.657	0.000000
Temperature	0.3525	1	0.3525	2.8020	0.125088
Time 2	0.0433	1	0.0433	0.3442	0.570427
Time 4	0.1583	1	0.1583	1.2583	0.288179
Time 6	0.7030	1	0.7030	5.5882	0.039685
Error	1.2580	10	0.1258		
Time	0.4889	2	0.2444	5.4270	0.013110
Time × Temperature	0.5521	2	0.2760	6.1290	0.008394
27°C	0.0010	2	0.0005	0.0111	0.988964
32°C	1.0400	2	0.5200	11.5479	0.000463
Error	0.9008	20	0.0450		
Acrylate in <i>Acropora tenuis</i>					
Effect	SS	df	MS	F	<i>p</i>
Intercept	903.6474	1	903.6474	589.2514	0.000000
Temperature	94.6001	1	94.6001	61.6869	0.000014
Time 2	0.6782	1	0.6782	0.4422	0.521095
Time 4	1.8205	1	1.8205	1.1871	0.301475
Time 6	3.0288	1	3.0288	1.9750	0.190211
Error	15.3355	10	1.5355		
Time	2.8944	2	1.4472	1.1180	0.346530
Time × Temperature	5.6668	2	2.8334	2.1888	0.138160
27°C	0.0105	2	0.0052	0.0041	0.995953
32°C	0.6705	2	0.3353	0.2590	0.774379
Error	25.8897	20	1.2945		

DMSP concentrations (**a**) and acrylate concentrations (**b**) are shown. Red lines show significant differences ($P < 0.05$).

Extended Data Table 3 | Sums of square (SS), mean squares (MS) and significance levels for ANOVAs of the thermal stress experiment on adult *Acropora millepora* corals

a	Effect	SS	df	MS	F	<i>p</i>
	Intercept	1308.7770	1	1308.7770	2314.789	0.000000
	Temperature	11.0560	1	11.0560	19.5550	0.000215
	Time -7	0.0105	1	0.0105	0.0186	0.893047
	Time 0	1.9423	1	1.9423	3.4377	0.077184
	Time 5	6.1001	1	6.1001	10.7966	0.003374
	Time 10	7.2047	1	7.2047	12.7517	0.001707
	Error	12.4390	22	0.5650		
	Time	8.7870	3	2.9290	11.2720	0.000005
	Time × Temperature	4.2010	3	1.4000	5.3890	0.002226
	27°C	1.5627	3	0.5209	2.0035	0.122017
	32°C	11.4258	3	3.8086	14.6485	0.000000
	Error	17.1510	66	0.2600		

b	Effect	SS	df	MS	F	<i>p</i>
	Intercept	3531.6680	1	3531.6680	3456.576	0.000000
	Temperature	14.4190	1	14.4190	14.1120	0.001090
	Time -7	0.0650	1	0.0650	0.0636	0.803235
	Time 0	3.0065	1	3.0065	2.9418	0.100373
	Time 5	8.5848	1	8.5848	8.4000	0.008338
	Time 10	10.1534	1	10.1534	9.9348	0.004626
	Error	22.4780	22	1.0220		
	Time	11.7760	3	3.9250	7.2940	0.000269
	Time × Temperature	7.3910	3	2.4640	4.5780	0.005668
	27°C	1.2300	3	0.4100	0.7621	0.519419
	32°C	17.9380	3	5.9790	11.1134	0.000005
	Error	35.5210	66	0.5380		

c	Effect	SS	df	MS	F	<i>p</i>
	Intercept	59.3872	1	59.3872	38632.320	0.000000
	Temperature	0.1444	1	0.1444	93.9100	0.000000
	Time -7	0.0006	1	0.0006	0.3896	0.538929
	Time 0	0.0020	1	0.0020	1.3247	0.262124
	Time 5	0.0223	1	0.0223	14.4545	0.000977
	Time 10	0.2928	1	0.2928	190.1039	0.000000
	Error	0.0338	22	0.0015		
	Time	0.1689	3	0.0563	103.2300	0.000000
	Time × Temperature	0.1733	3	0.0578	105.9100	0.000000
	27°C	0.0005	3	0.0002	0.3273	0.805680
	32°C	0.3417	3	0.1139	207.0727	0.000000
	Error	0.0360	66	0.0005		

DMSP concentrations (a), acrylate concentrations (b) and PAM data (c) are shown. Red lines show significant differences ($P < 0.05$).

Extended Data Table 4 | Description of the reductase and methyltransferase sequences in diatoms, corals and *Symbiodinium*

	Reductase	Methyltransferase
<i>Fragilariopsis cylindrus</i>	jgi Fracy1 173405 estExt _Genewise1.C_220021	jgi Fracy1 207357 estExt _Genewise1Plus.C_41074
<i>Acropora digitifera</i>	adi_v1.10730	adi_v1.21031
<i>Acropora millepora</i>	Cluster027405	Cluster022229
<i>Symbiodinium</i> (clade A, strain k8)	kb8_c41244	kb8_rep_c2522
Pfam domain	FMN_red (PF03358.8)	Methyltransf_7 (PF03492.8)
OrthoMCL cluster	OG5_131390	OG5_156314

Extended Data Table 5 | Composition of coral juvenile bacterial communities present in the two different coral species at the two temperature regimes after 6 days

Genus	<i>A. tenuis</i> 27°C	<i>A. tenuis</i> 32°C	<i>A. millepora</i> 27°C	<i>A. millepora</i> 32°C	DMSP degradation	DMS degradation
<i>Alteromonas</i>		16.7	6.3	2.1	43	
<i>Endozoicomonas</i>			8.3	10.4	43	
<i>Hyphomonas</i>	4.2		2.1	2.1	44	
<i>Idiomarina</i>	2.1				43	
<i>Labrenzia</i>				4.2	45	
<i>Marinomonas</i>		2.1		4.2	46	
<i>Neptuniibacter</i>	22.9	4.2	8.3	2.1	47	
<i>Oceanicola</i>		2.1	2.1	2.1	48	
<i>Phaeobacter</i>	22.9	4.2	12.5	6.3	49	
<i>Pseudoalteromonas</i>			2.1		44	
<i>Ruegeria</i>	25.0	29.2	33.3	20.8	50	
<i>Sagittula</i>				2.1	46	
<i>Stappia</i>				2.1	48	
<i>Methylophaga</i>	2.1					51
<i>Rhodovulum</i>		12.5	10.4	25.0		52
<i>Bdellovibrio</i>	2.1					
<i>Hellea</i>	2.1					
<i>Krokinobacter</i>	2.1					
<i>Lacinutrix</i>	2.1					
<i>Maricaulis</i>	2.1		2.1			
<i>Mesorhizobium</i>				4.2		
<i>Methylosinus</i>	2.1					
<i>Planifilum</i>			2.1			
<i>Ponticaulis</i>	2.1	10.4	2.1	4.2		
<i>Pseudovibrio</i>		2.4				
<i>Ralstonia</i>	2.1			2.1		
<i>Shimia</i>		6.3	4.2	4.2		
<i>Simiduia</i>		2.1				
<i>Spongiibacter</i>	4.2	4.2	2.1	2.1		
<i>Thalassospira</i>		4.2	2.1			

Data are presented at the genus level in per cent; 48 clones per sample. Bacteria previously implicated in the degradation of DMSP and DMS are shown in green and orange, respectively. Note the proportion of bacteria involved in DMS(P) degradation, representing 79.7% of the bacteria sequenced. References ⁴³⁻⁵² are cited in this table.

Structural insight into magnetochrome-mediated magnetite biomineralization

Marina I. Siponen^{1,2,3}, Pierre Legrand⁴, Marc Widdrat⁵, Stephanie R. Jones⁶, Wei-Jia Zhang^{7,8†}, Michelle C. Y. Chang⁶, Damien Faivre⁵, Pascal Arnoux^{1,2,3,8} & David Pignol^{1,2,3,8}

Magnetotactic bacteria align along the Earth's magnetic field using an organelle called the magnetosome, a biomineralized magnetite (Fe(II)Fe(III)₂O₄) or greigite (Fe(II)Fe(III)₂S₄) crystal embedded in a lipid vesicle. Although the need for both iron(II) and iron(III) is clear, little is known about the biological mechanisms controlling their ratio¹. Here we present the structure of the magnetosome-associated protein MamP and find that it is built on a unique arrangement of a self-plugged PDZ domain fused to two magnetochrome domains, defining a new class of *c*-type cytochrome exclusively found in magnetotactic bacteria. Mutational analysis, enzyme kinetics, co-crystallization with iron(II) and an *in vitro* MamP-assisted magnetite production assay establish MamP as an iron oxidase that contributes to the formation of iron(III) ferrihydrite eventually required for magnetite crystal growth *in vivo*. These results demonstrate the molecular mechanisms of iron management taking place inside the magnetosome and highlight the role of magnetochrome in iron biomineralization.

Magnetotactic bacteria (MTB) have the particular ability to align with geomagnetic field lines, a phenomenon referred to as magnetotaxis. This magnetotactic property is due to the presence of the magnetosome, an organelle made of a lipid vesicle loaded with a single magnetite or greigite crystal about 50 nm in size. The alignment of magnetosomes inside the cell acts like a compass needle to orient MTB passively in geomagnetic fields, putatively simplifying their search for preferred microaerophilic environments. Formation of this iron-rich organelle is genetically orchestrated by genes located in the magnetosome genetic island. These genes ensure the formation of the vesicles, their alignment, their loading with iron and the biomineralization into magnetite or greigite^{2,3}. Despite early observations of redox control in MTB^{4,5}, this last step remains poorly understood, notably the management of the iron(II) and iron(III) species required for magnetite or greigite formation. There is indication that some oxidized iron species such as ferrihydrite accumulate before magnetite formation, therefore suggesting the need for a reductive process^{6,7}. However, there is also growing evidence that the readily available iron species in the magnetosome is iron(II). Both the presence of numerous and active ferric reductases in MTB⁸ and the predominance of cation diffusion facilitators for iron(II) trafficking⁹ associated with the magnetosome support this premise.

The search for potential redox proteins within the magnetosome genetic island has led to the identification of four gene products containing at least two tandem *c*-type cytochrome motifs CX₂CH, recently called magnetochrome domains¹⁰: MamE, MamP, MamT and MamX. Among these magnetochrome-containing proteins, MamE and MamP are conserved in all MTB and, interestingly, deletion mutants of the corresponding genes show defects in the biocrystallization process^{11,12}. However, the multiplicity of this domain leads to difficulties in the phenotypic analyses of a single-domain deletion mutant in the *mamE* gene¹³, suggesting that

magnetochrome domains could be functionally redundant. This magnetochrome domain seems specific to MTB as it has not been found in any other species so far, suggesting it may represent a new functional class of cytochrome.

We purified and crystallized the soluble part of MamP from the MO-1 strain (residues 26–260, see Methods). The MamP structure was subsequently solved by multi-wavelength anomalous diffraction (MAD) using the four iron atoms present in the asymmetric unit (see Extended Data Table 1 for statistics and Extended Data Fig. 1 for an example of the $2mF_{\text{obs}} - DF_{\text{calc}}$ map). The first visible residue in the electron density corresponds to residue 87, indicating the presence of a long flexible arm connecting the protein to the single transmembrane helix. Following this flexible arm, the protein folds as a PDZ domain, a small *c*-type cytochrome domain (the first magnetochrome domain, MCR1), a 17-residue linker and finally a second magnetochrome domain (Fig. 1; see Extended Data Fig. 2 for an annotated sequence alignment). The first magnetochrome domain is in contact with its own PDZ domain, whereas the second is projected above the PDZ domain of the other monomer. The minimal unit of MamP is a dimer, although the crystal packing could also support the existence of a tetramer made by two symmetric dimers (one in an 'open' state and the other in a 'closed' state; Extended Data Figs 3 and 4) differing by only small but functionally important side-chain reorientations, as outlined below. In solution, size exclusion chromatography indicates a pH-dependent tetramer/dimer equilibrium that was confirmed by small-angle X-ray scattering (SAXS) whereas circular dichroism measurements indicated no major structural rearrangement upon pH change (Methods and Extended Data Fig. 3).

Using the PDZ domain of MamP, a structural homology search using the DALI server¹⁴ indicated that the closest structural homologues are the PDZ domains found in the high-temperature requirement A (HtrA) family of Ser proteases. HtrA proteases combine a protease domain to one or two PDZ domains and are involved in protein quality control^{15,16}. These functions are made possible by the peptide-binding properties of PDZ domains, a domain that folds as a single β -sheet capped on one side by an α -helix, thereby delineating a groove dedicated to peptide substrates by β -strand augmentation¹⁷.

Interestingly, the PDZ domain of MamP is unusual because its groove is not open for protein partner binding. Instead, the first visible strand in the MamP structure (β 1, denoted as SP for 'self-plugging' strand) fills the binding groove found in classical PDZ domains. In MamP, this SP strand clearly contributes to its dimerization (Fig. 2a). Indeed, the position of the SP strand further allows the extension of the β -sheet, with strand β 7 connecting the PDZ domain to the first magnetochrome domain. This last strand (denoted as Dim for dimerization strand) largely contributes to the dimeric interface of MamP (Fig. 2a). The overall

¹Commissariat à l'Energie Atomique et aux Energies Alternatives, Direction des Sciences du Vivant, Institut de Biologie Environnementale et de Biotechnologies, Laboratoire de Bioénergétique Cellulaire, Saint-Paul-lez-Durance, F-13108, France. ²Centre National de la Recherche Scientifique, Unité Mixte de Recherche Biologie Végétale et Microbiologie Environnementales, Saint-Paul-lez-Durance, F-13108, France. ³Aix-Marseille Université, Saint-Paul-lez-Durance, F-13108, France. ⁴Synchrotron SOLEIL, L'Orme des Merisiers Saint-Aubin, 91192 Gif-sur-Yvette, France. ⁵Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Science Park Golm, 14424 Potsdam, Germany. ⁶Departments of Chemistry and Molecular and Cell Biology, University of California, Berkeley, California 94720-1460, USA. ⁷Aix-Marseille Université, Laboratoire de Chimie Bactérienne, UMR7283, Institut de Microbiologie de la Méditerranée, CNRS, F-13402 Marseille Cedex 20, France. ⁸Laboratoire International Associé Biominéralisation et Nanostructure, CNRS-Marseille, F-13402 Marseille Cedex 20, France. [†]Present address: MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

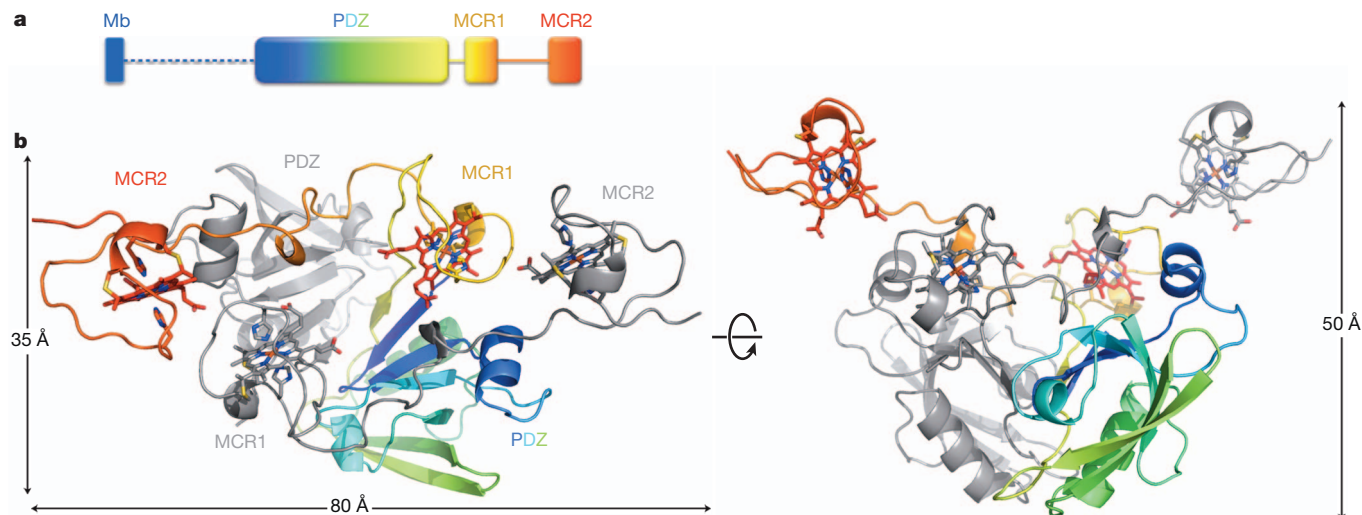


Figure 1 | Overall structure of MamP homodimer. **a**, Representation of MamP domain organization. The linker between the transmembrane helix and the PDZ domain is shown as a dashed line, indicating that it is disordered in the

crystal structure. **b**, Three-dimensional structure of MamP with one monomer coloured in grey and the other monomer coloured with a ramp from blue (amino (N) terminus) to red (carboxy (C) terminus).

interface is conserved, suggesting a selective pressure for this oligomeric assembly.

We recently proposed that the two *c*-type cytochrome domains of MamP define a new domain exclusively found in MTB¹⁰. The present structure determination of MamP allows us to describe the fold of this domain, demonstrating its uniqueness at the structural level. A magnetochrome starts with a hydrophobic residue ($\psi 1$ in Fig. 2b and Extended Data Fig. 2) in direct contact with the haem, followed by a Pro-His (PH) dyad, located five to nine residues upstream of the CXXCH motif, providing the sixth and fifth haem ligands, respectively. Finally, a terminal hydrophobic residue ($\psi 2$ in Fig. 2b) closes the magnetochrome fold through a hydrophobic interaction with the $\psi 1$ residue.

The structure of MamP also confirms that the magnetochrome domain defines a single haem-binding domain belonging to a new family of *c*-type cytochrome. Indeed, it folds as one of the smallest haem-binding units known thus far, with only 23 residues surrounding the haem (Fig. 2b). This best compares to artificial microperoxidases that possess a covalently attached haem and a single histidine ligand, whereas other mono-haem *c*-type cytochromes minimally possess about 70 residues surrounding

a single haem (Protein Data Bank accession number 1K3G). We found that the haems in both magnetochrome domains are highly solvent exposed, with 281 and 214 Å² for MCR1 and MCR2, respectively. These values best compare to multihaem cytochromes or proteins with transient affinity for haem such as haemophores or haemopexin¹⁸. In addition, the haem-binding mode in magnetochrome domains also stands out because all rings of the haems are solvent exposed, which is not the case in other *c*-type cytochromes¹⁸.

The dimeric structure of MamP creates a large surface-exposed acidic pocket resembling a crucible with approximate dimensions of 8 Å (depth) \times 15 Å (diameter) (Fig. 3a). Eight conserved acidic residues from both PDZ domains delineate the bottom of this crucible, whereas the sides are formed by the propionates of the haems and four conserved acidic residues from the linkers between the two magnetochrome domains (Fig. 3a). A conserved histidine residue (H93) is located in the middle of the crucible with a network of conserved hydrogen bonded residues connecting its side chain to the exterior of the protein through polar residues, which is reminiscent of a hydrogen exit channel (Fig. 3b and Extended Data Fig. 5). The peculiar arrangement of conserved acidic residues observed in the MamP cavity suggests the presence of a 'hot spot' and led us to investigate how it reacts with iron compounds. *In vitro*, we found that MamP efficiently oxidized Fe(II)SO₄ at alkaline pH (Fig. 4a). This reaction proceeded with a rate constant (k_{ox}) of $1.06 \times 10^{-3} \mu\text{M}^{-1} \text{s}^{-1}$ at an optimal pH of 9, which is comparable to observations on the multihaem cytochrome *c* MtoA, a decahaem *c*-type cytochrome from *Sideroxydans lithotrophicus* involved in microbial iron oxidation¹⁹. Interestingly, the optimal pH of MamP iron oxidase activity coincides with that found for *in vitro* magnetite synthesis²⁰.

The Fe(II) oxidation activity detected with MamP *in vitro* further prompted us to investigate whether this activity could be substantiated by structural approaches and by testing the effect of MamP on magnetite formation in solution. To get an impression of the iron oxidase activity of MamP, we transferred a MamP crystal to pH 9 and then soaked it in a solution containing Fe(II)SO₄ before data collection at the iron edge. The resulting anomalous electron density clearly indicates the presence of an iron-binding site mediated by conserved residues located at the bottom of the crucible (Fig. 4b, c). The anomalous electron density peak is only present in the open dimer and its elongated shape even suggests the presence of a di-iron-binding site with the two iron atoms replacing the two water molecules that are only seen in the high-resolution structure of the open dimer. This stabilization of two iron atoms in the open dimer is in line with the calculated charged

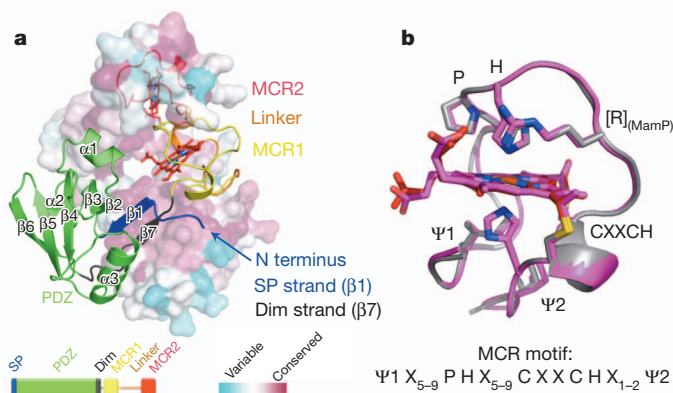


Figure 2 | A conserved dimerization interface mediated by the PDZ domain and structure of a magnetochrome domain. **a**, The surface representation of the one monomer is colour-coded based on sequence conservation in alignments of all known MamP and indicates that the dimeric interface is conserved. Note that both the self-plugged (SP) and the dimerization (Dim) strands participate in the dimerization of MamP together with strand $\beta 3$ and helices $\alpha 1$ and $\alpha 2$. **b**, Superimposition of MCR1 with MCR2 with conserved magnetochrome residues represented in stick.

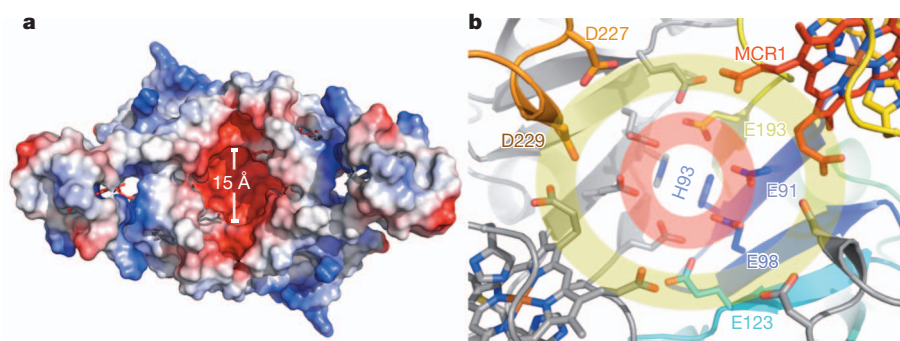


Figure 3 | A crucible on the surface of MamP is built on a conserved acidic pocket surrounded by a conserved acidic crown. **a**, Molecular surface representation of a MamP dimer coloured according to its electrostatic

states, which suggests that the iron-binding residues are unprotonated in the open dimer (Extended Data Fig. 4). Using a ferrozine assay to estimate the iron/MamP stoichiometry, we found that four irons are oxidized per MamP dimer, which provides a better fit with a di-iron-binding site.

To examine the functional relevance of the conserved acidic residues, some of which are directly involved in iron binding, we used the *mamP* deletion mutant of *Magnetospirillum magneticum* AMB-1 (the genetic tools being available for this strain), and complemented this strain with the wild-type gene or a *mamP* variant in which all conserved acidic residues are mutated to alanine (*mamP* Δ_{acid}). In accordance with our observation of an iron-binding site at the bottom of the crucible, we found that these residues are essential for magnetite formation *in vivo* as judged by the magnetic response of the cells (C_{mag}), crystal size

potential. The size of the crucible is indicated. **b**, Detail of the two acidic networks making the bottom (red circle) and the side (yellow circle) of the crucible.

distributions as well as transmission electron microscope (TEM) images (Fig. 4d, e and Extended Data Figs 6 and 7).

Finally, we observed the role of MamP in a mineralization experiment (Fig. 4f and Extended Data Figs 8 and 9). Indeed, magnetite is typically formed in solution by co-precipitation experiments of iron with the stoichiometric ratio of magnetite (Fe(II)/Fe(III) = 0.5) (ref. 21). We decided to start exclusively with Fe(II) to see if any mineral would form in the presence or absence of MamP. In the presence of MamP, we initially observed the formation of ferrihydrite, an Fe(III) oxide, with a progressive evolution of this mineral to magnetite (Fig. 4f). The control experiment omitting MamP could not allow the detection of any mineral by X-ray diffraction. TEM images indicated the presence of electron-dense particles that were not seen when MamP was omitted, thereby suggesting MamP-mediated production of ferrihydrite or magnetite

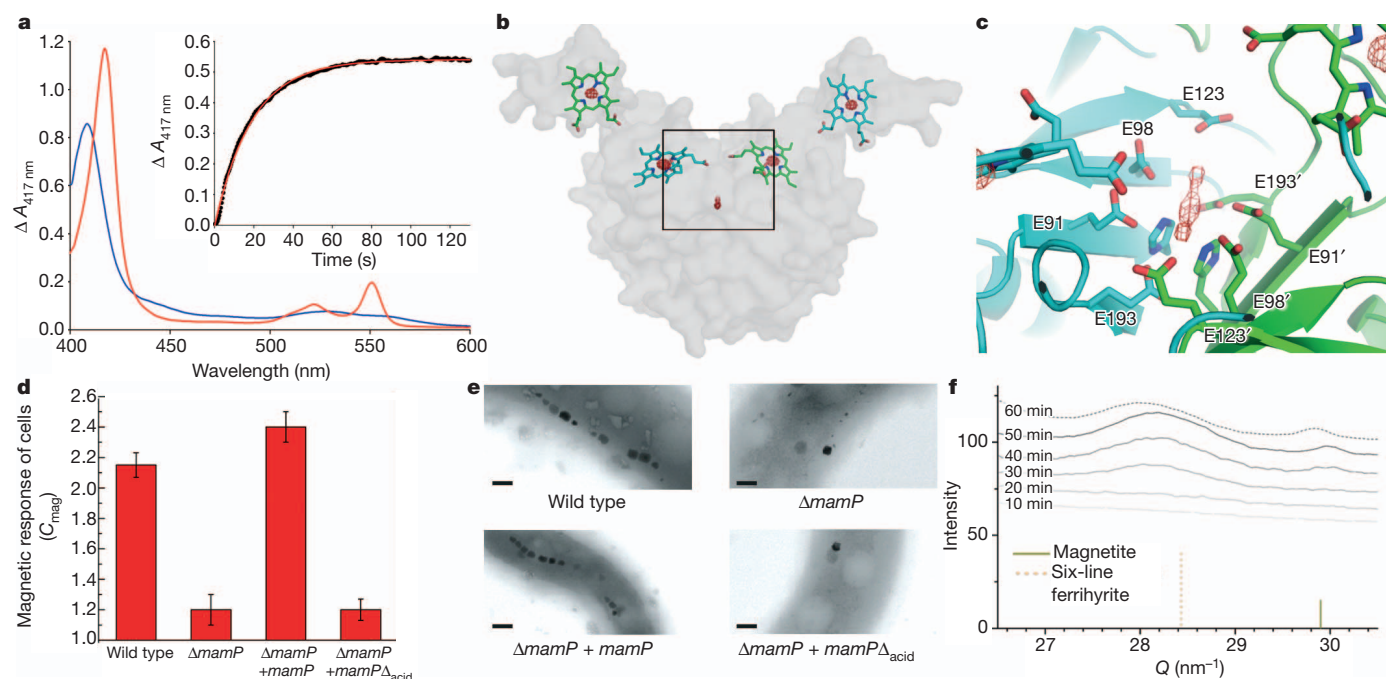


Figure 4 | MamP, an iron oxidase with a functionally important crucible, mediates ferrihydrite production in an *in vitro* mineralization experiment starting with Fe(II). **a**, Visible spectra of MamP (5 μ M) before (blue line) and after (red line) addition of 50 μ M Fe(II)SO₄. Inset: kinetics of reduction of MamP by Fe(II)SO₄ (black dots) and fit of the experimental data points by a single exponential (red line). **b**, Soaking experiments of MamP crystals with Fe(II)SO₄ at pH 9. The electron density map corresponds to an anomalous map collected at the iron edge and contoured at 5 σ all around the dimer. **c**, Details of the position of the anomalous peak in the conserved acidic pocket. **d**, **e**, Genetic complementation studies to examine the function of MamP acidic residues.

d, Magnetic response of cells. Data were collected as three biological replicates with two technical replicates per biological replicate. Values are represented as mean \pm s.d. ($n = 6$). **e**, Representative TEM images of strains (scale bar, 0.2 μ m). Cell images are representative of those collected for all three replicates. **f**, Time-resolved analysis of the mineralization synthesis followed by X-ray diffraction with reference peak of ferrihydrite and magnetite and their relative intensity (the full X-ray diffraction spectrum is shown in Extended Data Fig. 7). The six-line ferrihydrite peak ((112) peak in red dotted line) increases until 50 min whereas the magnetite peak ((400) in green) is only visible after 40 min and increases then, possibly at the expense of the ferrihydrite precursor.

(Extended Data Fig. 8). This mechanism is in accord with a process where the Fe(II) is oxidized by the protein to enable the formation of ferrihydrite, a purely ferric iron oxide. Once MamP is fully reduced, the continuous addition of Fe(II) enables the transformation of ferrihydrite to magnetite. Such a mechanism strongly resembles the pathway recently described for the synthetic formation of magnetite from solution²².

The structural basis of MamP-mediated iron biomineralization presented here validates an old model put forth for how magnetite formation should be redox-controlled in the magnetosome, and emphasizes the versatility of the magnetochrome domains in this process⁵. Magnetite crystal growth requires a precise Fe(III)/Fe(II) ratio; the MamP properties we showed, both *in vitro* and *in crystallo*, would allow the production of the ferrihydrite precursor thanks to the presence of acidic residues at the bottom of a crucible surrounded by four magnetochrome domains. Together with the presence of iron(II), this ferrihydrite would evolve towards magnetite in the magnetosome. This molecular model fits perfectly with sequential events observed for magnetite biomineralization in MTB^{6,7}. The dimension and acidic nature of the MamP crucible, the presence of a conserved proton exit channels at its bottom and the four haems on either side are well-suited for the expected chemistry of ferrihydrite formation from Fe(II):



Functional and structural studies are now required to determine the contribution of other magnetochrome-containing proteins in the process of magnetite biomineralization.

METHODS SUMMARY

MamP from strain MO-1 was cloned in plasmid pET26b+ and expressed in *Escherichia coli*. The protein was purified by metal-affinity and gel-filtration chromatography. Protein crystals were obtained at 20 °C, diffraction data were collected at synchrotrons SOLEIL, SLS and ESRF, and the structure was solved by MAD. Iron oxidase activity was measured in an anaerobic glove box. We used strain AMB-1 for *in vivo* mutational analysis. *In vitro* biomineralization experiments were done under nitrogenous atmosphere using Fe(II), and X-ray diffraction was measured at synchrotron BESSY II.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 April; accepted 13 August 2013.

Published online 6 October; corrected online 30 October 2013 (see full-text HTML version for details).

1. Blakemore, R. Magnetotactic bacteria. *Science* **190**, 377–379 (1975).
2. Schuler, D. Genetics and cell biology of magnetosome formation in magnetotactic bacteria. *FEMS Microbiol. Rev.* **32**, 654–672 (2008).
3. Komeili, A. Molecular mechanisms of compartmentalization and biomineralization in magnetotactic bacteria. *FEMS Microbiol. Rev.* **36**, 232–255 (2012).
4. Bell, P. E., Mills, A. L. & Herman, J. S. Biogeochemical conditions favoring magnetite formation during anaerobic iron reduction. *Appl. Environ. Microbiol.* **53**, 2610–2616 (1987).
5. Frankel, R. B. & Blakemore, R. P. Precipitation of Fe₃O₄ in magnetotactic bacteria. *Phil. Trans. R. Soc. Lond. B* **304**, 567–573 (1984).
6. Baumgartner, J. *et al.* Magnetotactic bacteria form magnetite from a phosphate-rich ferric hydroxide via nanometric ferric (hydr)oxide intermediates. *Proc. Natl Acad. Sci. USA* **110**, 14883–14888 (2013).

7. Fdez-Gubieda, M. L. *et al.* Magnetite biomineralization in *Magnetospirillum gryphiswaldense*: time-resolved magnetic and structural studies. *ACS Nano* **7**, 3297–3305 (2013).
8. Zhang, C. *et al.* Two bifunctional enzymes with ferric reduction ability play complementary roles during magnetosome synthesis in *Magnetospirillum gryphiswaldense* MSR-1. *J. Bacteriol.* **195**, 876–885 (2012).
9. Uebe, R. *et al.* The cation diffusion facilitator proteins MamB and MamM of *Magnetospirillum gryphiswaldense* have distinct and complex functions, and are involved in magnetite biomineralization and magnetosome membrane assembly. *Mol. Microbiol.* **82**, 818–835 (2011).
10. Siponen, M. I., Adryanczyk, G., Ginet, N., Arnoux, P. & Pignol, D. Magnetochrome: a c-type cytochrome domain specific to magnetotactic bacteria. *Biochem. Soc. Trans.* **40**, 1319–1323 (2012).
11. Lohsse, A. *et al.* Functional analysis of the magnetosome island in *Magnetospirillum gryphiswaldense*: the mamAB operon is sufficient for magnetite biomineralization. *PLoS ONE* **6**, e25561 (2011).
12. Murat, D., Quinlan, A., Vall, H. & Komeili, A. Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proc. Natl Acad. Sci. USA* **107**, 5593–5598 (2010).
13. Quinlan, A., Murat, D., Vall, H. & Komeili, A. The HtrA/DegP family protease MamE is a bifunctional protein with roles in magnetosome protein localization and magnetite biomineralization. *Mol. Microbiol.* **80**, 1075–1087 (2011).
14. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
15. Clausen, T., Kaiser, M., Huber, R. & Ehrmann, M. HTRA proteases: regulated proteolysis in protein quality control. *Nature Rev. Mol. Cell Biol.* **12**, 152–162 (2011).
16. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**, 455–459 (2002).
17. Lee, H. J. & Zheng, J. J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun. Signal.* **8**, 8 (2010).
18. Smith, L. J., Kahraman, A. & Thornton, J. M. Heme proteins—diversity in structural characteristics, function, and folding. *Proteins* **78**, 2349–2368 (2010).
19. Liu, J. *et al.* Identification and characterization of MtoA: a decaheme c-type cytochrome of the neutrophilic Fe(II)-oxidizing bacterium *Sideroxydans lithotrophicus* ES-1. *Front. Microbiol.* **3**, 37 (2012).
20. Baumgartner, J., Bertinetti, L., Widdrat, M., Hirt, A. M. & Faivre, D. Formation of magnetite nanoparticles at low temperature: from superparamagnetic to stable single domain particles. *PLoS ONE* **8**, 3 (2013).
21. Jolivet, J. P., Chaneac, C. & Tronc, E. Iron oxide chemistry. From molecular clusters to extended solid networks. *Chem. Commun. (Camb.)* **5**, 481–487 (2004).
22. Baumgartner, J. *et al.* Nucleation and growth of magnetite from solution. *Nature Mater.* **12**, 310–314 (2013).

Acknowledgements This work received institutional support from the Commissariat à l’Energie Atomique et aux Energies Alternatives, the Centre National de la Recherche Scientifique, Aix-Marseille University and the Max Planck Society. We are grateful to BM-30 (ESRF, Grenoble, France) and X06SA (SLS, Villigen, Switzerland) staff for technical assistance in synchrotron data collection. We thank J. Perez (SOLEIL, GIF-sur-Yvette) for help in SAXS data collection, and A. Komeili for the gift of the wild-type and Δ mamP AMB-1 strains. We acknowledge S. Siegel and C. Li for their support at the μ Spot beamline of BESSY II, Helmholtz Zentrum Berlin. We thank the AFMB laboratory (Marseille) for circular dichroism measurements. M.I.S. was supported by a grant from the Eurotalent and ToxNuc-E programs. D.F. is supported by the Max Planck Society and a Starting Grant from the ERC (256915-MB2). S.R.J. and M.C.Y.C. thank the Defense Advanced Research Projects Agency (N66001-12-1-4230) for support.

Author Contributions M.I.S., M.W., S.R.J. and P.A. performed experiments. M.I.S., P.L. and P.A. performed structure determination. W.-J.Z. prepared genomic DNA. M.I.S., M.W., S.R.J., M.C.Y.C., D.F., P.A. and D.P. analysed the data. M.I.S., D.F., P.A. and D.P. prepared the manuscript. D.F., M.C.Y.C., P.A. and D.P. supervised the work.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 4JJ0 (apo form) and 4JJ3 (in complex with iron). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.P. (david.pignol@cea.fr) or P.A. (pascal.arnoux@cea.fr).

METHODS

Cloning, protein production and purification. The DNA sequence corresponding to residues D26–Q260 of magnetotactic ovoidal bacterium MO-1 *mamP* gene was sub-cloned into the plasmid pET26b+ (Novagen) using the In-Fusion PCR Cloning system (Clontech). Amplification was performed with *Pfu* Ultra High-Fidelity DNA polymerase (New England Biolabs). The resulting clone was co-expressed with the pEC26 (a gift from R. van Lis) clone of the *ccm* operon in chemically competent *E. coli* BL21(DE3) cells (Invitrogen).

Bacterial cells were cultured at 37 °C in Super Optimal Broth supplemented with 50 µg ml⁻¹ kanamycin and 50 µg ml⁻¹ chloramphenicol. Expression was induced once cells reach an absorbance of 0.6–0.8 by addition of 20 µM IPTG and left overnight at 37 °C. Cells were collected by centrifugation and re-suspended in lysis buffer (100 mM Na₂HPO₄, 500 mM NaCl, pH 8.0) supplemented with protease inhibitor cocktail (Sigma-Aldrich) and DNase. Lysis buffer was used in the ratio of 1.5 ml buffer per 1 g of wet cell pellet. Re-suspended pellets were stored at –80 °C.

Cells were disrupted by the One Shot cell disruption system (Constant Systems). Purification of the His-tagged proteins was performed in two steps using Ni-charged HiTrap Chelating HP and HiLoad 16/60 Superdex 200 columns on an ÄKTA FPLC protein purification system (GE Healthcare). Prep Grade columns (GE Healthcare) were pre-equilibrated with IMAC buffer (100 mM HEPES, 500 mM NaCl, pH 7.5) and gel filtration buffer (20 mM HEPES, 200 mM, pH 7.5), respectively. The filtered lysates were loaded onto the HiTrap Chelating columns and washed with IMAC buffer. Bound protein was eluted with IMAC buffer containing 100 mM imidazole and loaded onto the pre-equilibrated gel filtration column. Fractions containing the target proteins were pooled and concentrated using a VIVASPIN centrifugal filter device with a cut-off size of 10 kDa. Protein purity was confirmed by SDS–polyacrylamide gel electrophoresis and protein samples were flash frozen and stored at –80 °C.

Circular dichroism and SAXS measurements. Circular dichroism measurements used MamP (0.1 mg ml⁻¹) in 20 mM phosphate buffer. SAXS measurements were done on beamline SWING at SOLEIL (Paris). This beamline was equipped with a size-exclusion chromatography column (Superdex-200 3.3/30) coupled with ultra-violet detection and followed by online SAX measurements. Molecular masses of the eluted peaks were estimated using the MoW program²³.

Crystallization, data collection, structure determination and validation. Native crystals were obtained by the sitting drop vapour diffusion method in a 96-well plate. Half a microlitre of the protein solution (9.7 mg ml⁻¹) in GF buffer was mixed with 0.5 µl of well solution consisting of 0.2 M ammonium nitrate and 20% (w/v) PEG 3350. The plate was incubated at 20 °C and crystals were obtained after 12–14 days. Crystals were quickly transferred to cryogenic solutions containing well solution, 20% ethylene glycol or glycerol, and flash frozen in liquid nitrogen. To obtain structural data on Fe(II) binding to MamP, a crystal of the protein obtained in 0.2 M ammonium acetate and 20% (w/v) PEG3350 was transferred in the same mother liquor but containing 0.1 M bis–tris propane at pH9. Fe(II)SO₄ was then added to a final concentration of 10 mM for approximately 5 minutes after which the crystal was transferred to a cryogenic condition at pH 9 containing 20% MPD before flash freezing in liquid nitrogen. For the soaking experiments, all the solutions used were de-gassed.

Native data were collected to 1.8 Å on beamline X06SA at Swiss Light Source (Switzerland), whereas MAD data were collected to 2.5 Å on beamline PROXIMA1 at SOLEIL synchrotron facility (Gif-sur-Yvette, France). The crystal soaked in Fe(II) was collected to 2.8 Å on beamline FIP (ESRF, France). For the native crystals, data integration was performed in XDS and scaled with XSCALE (MAD data)²⁴ or SCALA (native data)²⁵. For the Fe(II)-soaked crystal, data integration was performed with Mosflm²⁶ and integrated with SCALA. The structure was solved by the autoSHARP program²⁷ using the MAD data set.

All the structures were refined with Refmac5 (ref. 28) with a final refinement cycle using PHENIX²⁹ in the case of the native structure. Structure validations were performed using the Molprobity server (<http://molprobity.biochem.duke.edu/>) and Procheck from the CCP4 suite.

Solution studies, mineralization experiments and analysis. The iron oxidase properties of MamP were studied in an anaerobic glove box filled with N₂ (<10 p.p.m. oxygen). The oxidation state of MamP was monitored by following the absorption of haem at either 417 or 551 nm. The ferrozine assay was also done in the anaerobic glove box. In this case, the diminution of Fe(II) concentration due to oxidation by MamP was estimated by mixing 400 µL of sample (5 µM MamP in a 50 mM bis–tris propane, 150 mM NaCl buffer, pH 8) with 1.6 ml of Ferrozine (1% w/v in 50 mM HEPES, pH 7). Fe(II) concentration was estimated by comparison with a standard curve using known Fe(II) concentrations.

The mineralization experiments were done similarly to the modified co-precipitation method developed by our group^{20,22}. Briefly, the computer-controlled setup consists of a titration device (Metrohm 888 Titrando), a dosing device (Metrohm 805 Dosimat) and a pH electrode (Metrohm Biotrode). The titration device provided a 0.1 M NaOH solution, and in the experiment reported here, the dosing device was filled with a 0.1 M ferrous solution to test the ability of the protein to oxidize iron to its ferric form and to the associated mineral phases. The solutions were prepared from 1 M sodium hydroxide solution (Merck) and ferrous chloride tetrahydrate (Sigma-Aldrich).

Before starting an experiment the original MamP-solution (3.2 ml, 200 µM; Hepes 20 mM, NaCl 500 mM) was dialysed with 200 mM NaCl solution to avoid the presence of foreign ions in the synthesis. All experiments were done at room temperature and under nitrogen atmosphere, the latter to avoid possible oxidation by air. All the solutions were carefully purged with nitrogen before use.

The reactor vessel was then filled with the dialysed MamP solution (14 ml, 22.9 µM) which was de-gassed for 30 min. Alternatively, the reaction vessel was filled with de-ionized water as a control. The pH was adjusted to 9 or 10 respectively before the crystallization experiment started. The pH was kept constant during the synthesis with the help of the titration device. The synthesis was performed by adding ferrous solution in the reactor through a micro capillary (Eppendorf microloader).

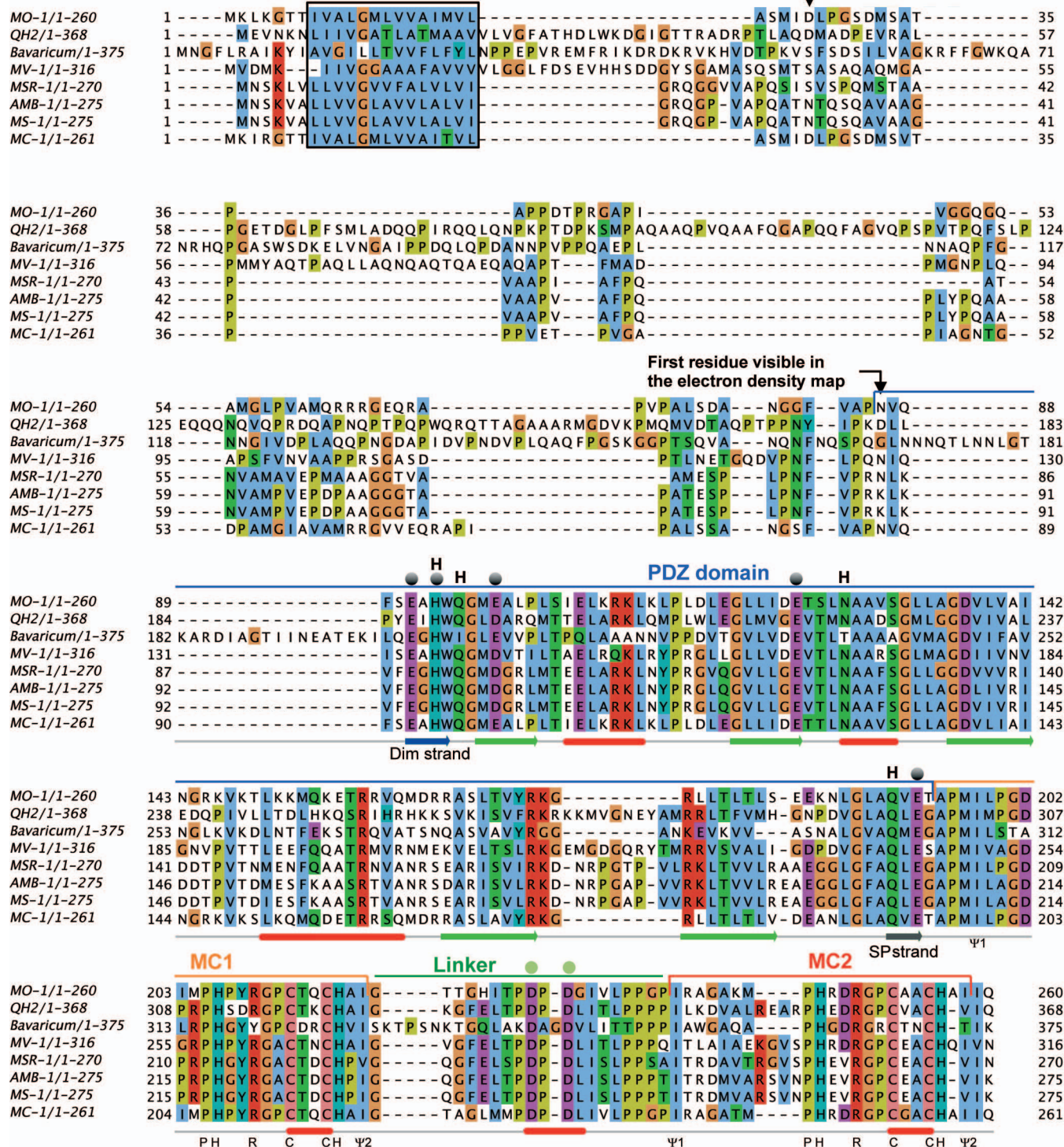
For imaging with electron microscopy, particles were adsorbed from the aqueous suspensions to the carbon film Cu grids. After removal of the liquid, the grids were washed with a drop of milli-Q water. The images were acquired with a Zeiss EM 912 Omega at 120 kV.

We used X-ray diffraction for mineralogical analyses. The samples were dried on a home-designed sample holder³⁰ and analysed. Because of the very reduced amount of samples, X-ray diffraction was measured at the µ-Spot synchrotron beamline (BESSY II, Helmholtz Zentrum Berlin) with a 100 µm beam of 15 keV (ref. 31).

23. Fischer, H., Neto, M., Napolitano, H. B., Craievich, A. F. & Polikarpov, I. The molecular weight of proteins in solution can be determined from a single SAXS measurement on a relative scale. *J. Appl. Cryst.* **43**, 101–109 (2010).
24. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010).
25. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr. D* **D62**, 72–82 (2006).
26. Leslie, A. G. W. & Powell, H. R. in *Evolving Methods for Macromolecular Crystallography* Vol. 245 (eds Read, R. J. & Sussman, J. L.) Ch. 4 41–51 (Springer, 2007).
27. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
28. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55**, 247–255 (1999).
29. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
30. Fischer, A., Schmitz, M., Aichmayer, B., Fratzl, P. & Faivre, D. Structural purity of magnetite nanoparticles in magnetotactic bacteria. *J. R. Soc. Interface* **8**, 1011–1018 (2011).
31. Paris, O. et al. A new experimental station for simultaneous X-ray microbeam scanning for small- and wide-angle scattering and fluorescence at BESSY II. *J. Appl. Cryst.* **40**, s466–s470 (2007).
32. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and interpretation of protein pKa values. *Proteins* **61**, 704–721 (2005).

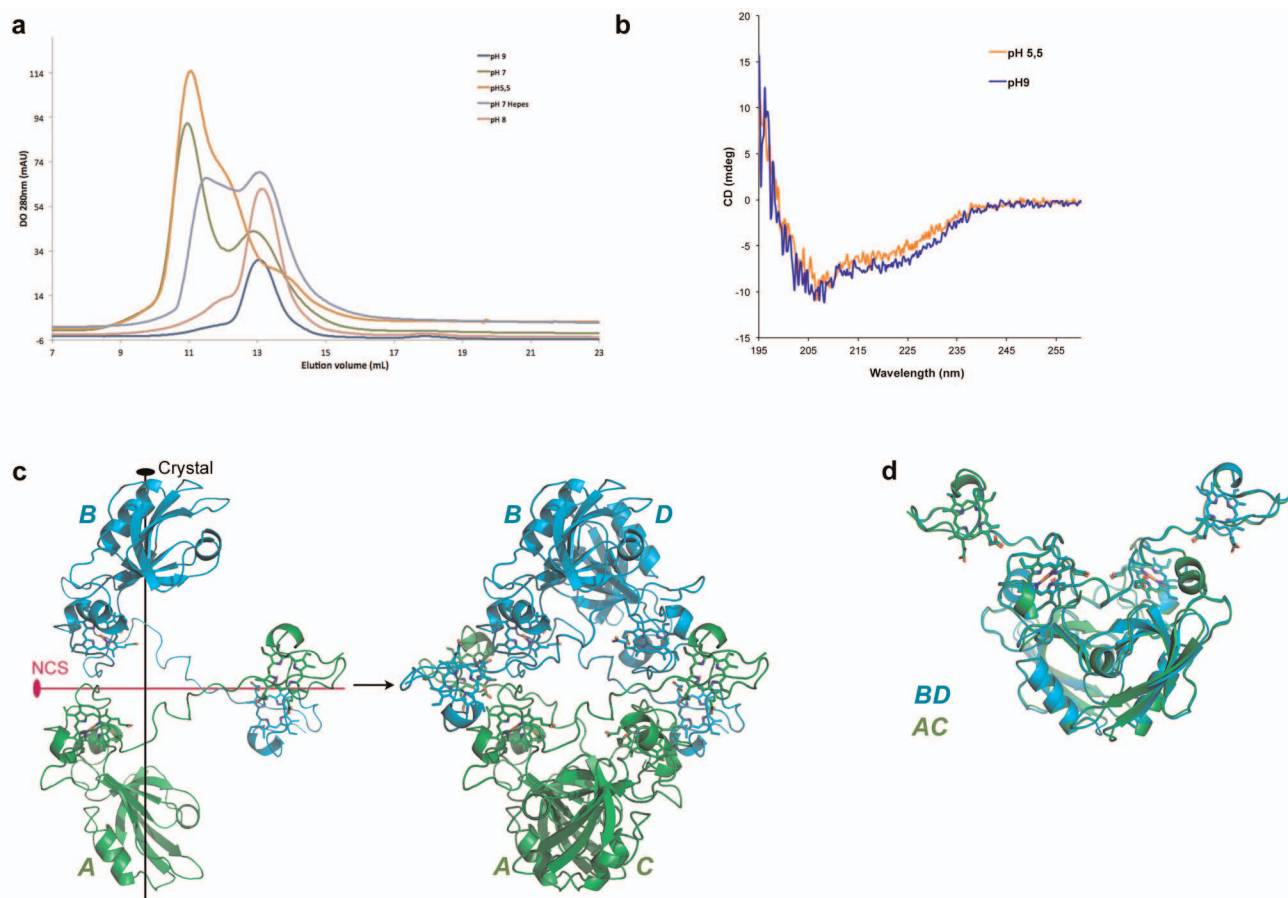
First residue of the cloned MamP

Predicted TM Helix



Extended Data Figure 2 | Sequence alignment of MamP proteins from different MTB and structural annotations discussed in the text. Black circles, acidic residues creating a hydrogen-bond network at the bottom of the crucible; green circles, acidic residues creating a hydrogen-bond network on the

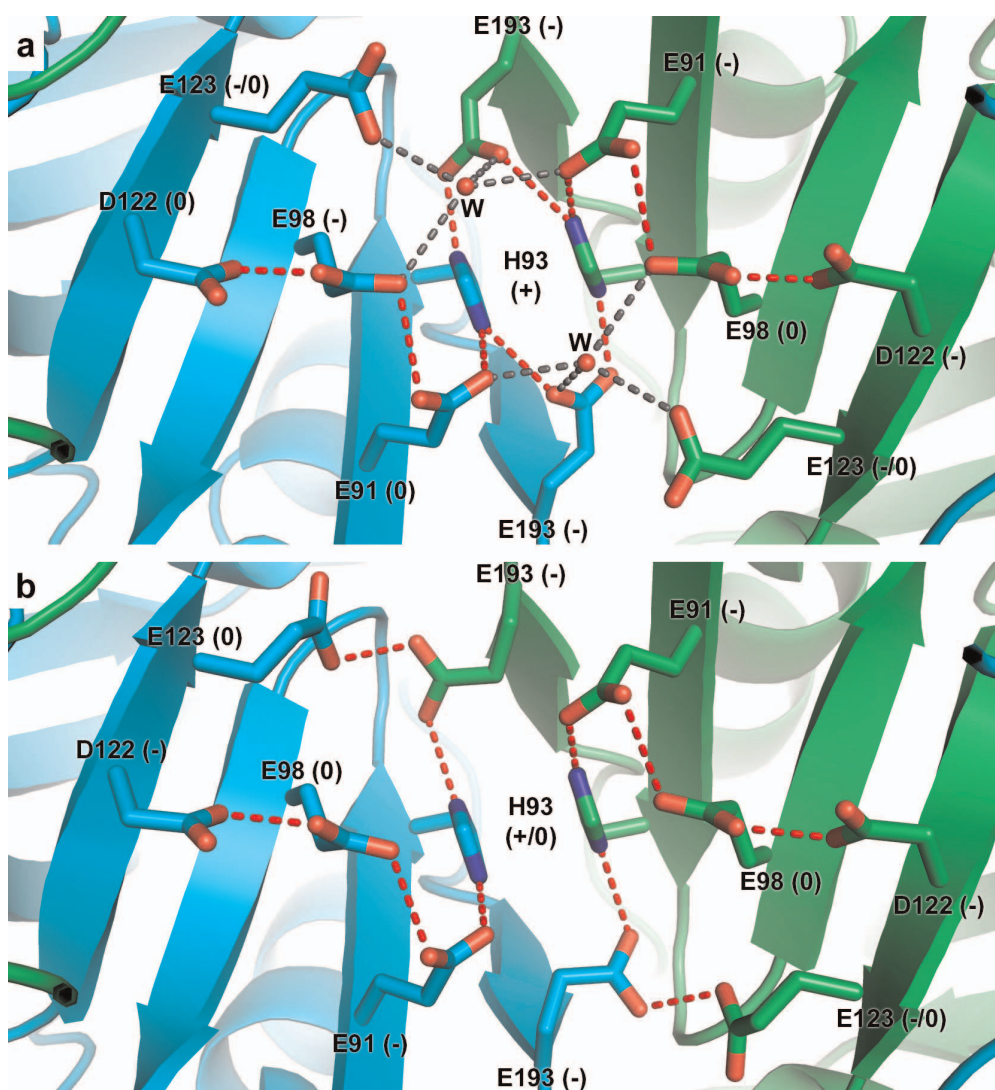
side of the crucible together with the propionate moieties of the haem from the MC1 domains; H, polar residues connecting H93 side chain to the exterior of the protein. Secondary structures are indicated at the bottom of the alignment.



Extended Data Figure 3 | pH-dependent oligomeric assembly of MamP.

a, Gel filtration of MamP using different buffers at different pH indicating a pH-dependent tetramer/dimer equilibrium. SAXS experiments confirm the presence of this equilibrium (see Methods). **b**, Circular dichroism measurement of MamP at pH 5 and 9 showing that there is no major structural rearrangement between the two pH values. **c**, The construction of the two different dimers of MamP (one in green, the other in cyan) starting from the two molecules in the asymmetric unit. These two molecules in the asymmetric unit are related by a non-crystallographic symmetry (NCS) represented in

magenta) axis. The two dimers (AC and BD) are generated using the twofold symmetry axis of the crystal (represented in black). The two dimers are therefore symmetric but they slightly differ, mainly in the orientation of two side chains of important residues located in the crucible (see Extended Data Fig. 4) supporting the notion of an 'open' (AC) and a 'closed' (BD) dimer. **d**, Superimposition of the two symmetric open and closed dimers. The root mean square distance between the C α positions of 176 superimposed residues is 0.51 Å, showing that there is no major structural difference between the two states.

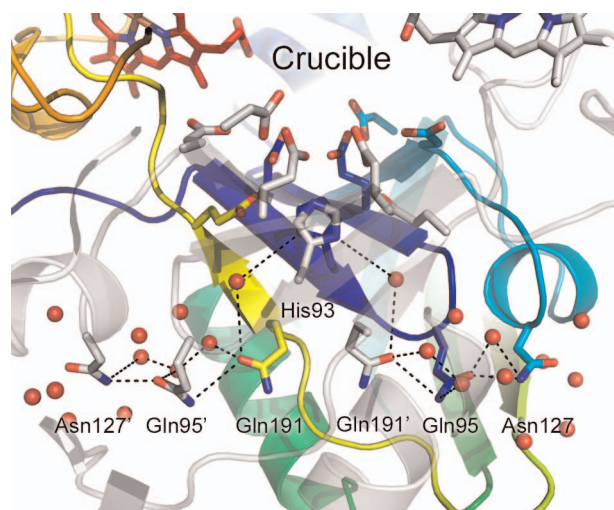


c

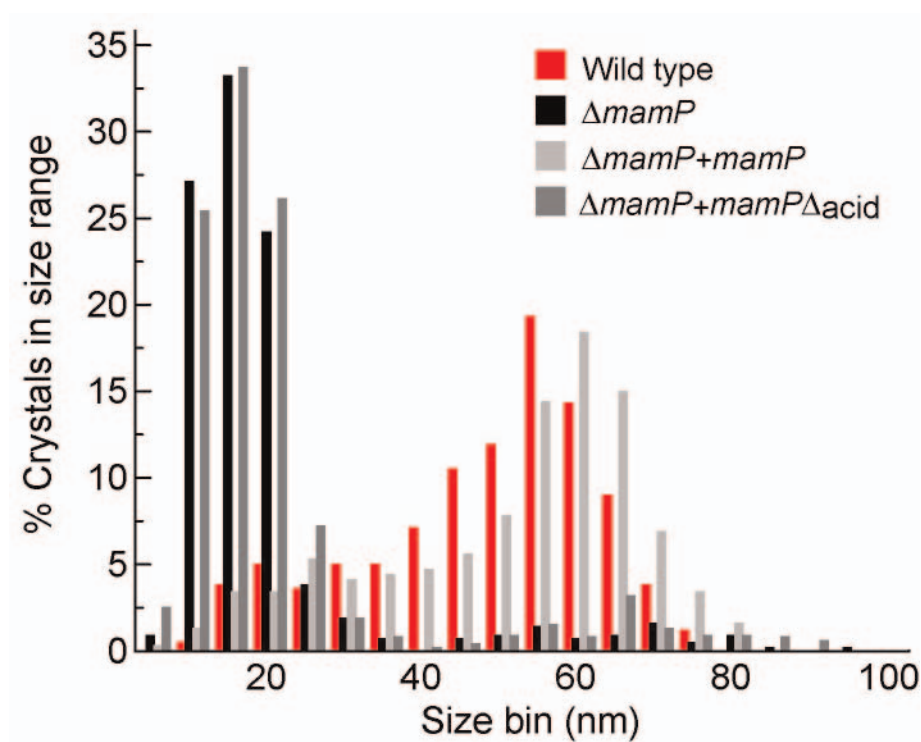
Residue	Open dimer		Closed Dimer	
	pKa (A)	pKa (C)	pKa (B)	pKa (D)
E91	10,6	7	2,2	2,7
H93	15,7	14,5	11,3	9,3
E98	4,3	16,3	15,5	15,3
E123	9,6	8,5	12,5	12,5
E193	0	0,5	4,7	4,7
D122	12,3	5,4	6,2	6,2
D227	3,8	3,8	3,6	3,6
D229	4,9	4,9	4,8	4,8

Extended Data Figure 4 | Putative hydrogen-bond network in the crucible of the two MamP dimers and protonation states at pH 9 deduced from pK_a calculations of protonable residues. a, Putative hydrogen-bond network and protonation states of the conserved acidic residues in the crucible of the AC (open) dimer of MamP. b, Putative hydrogen-bond network and protonation states in the BD (closed) dimer of MamP. Note the small reorientation of the side chains of E193 and E123 and the repercussion on the calculated charge and, ultimately, the stabilization of two water molecules at the dimeric interface: in the open dimer, the two side chains could stabilize two water molecules

(W) through two putative hydrogen bonds, which is not the case in the closed dimer. In the Fe(II) soaking experiment, the anomalous electron density extends towards these two water molecules in the open dimer, whereas it is not visible in the closed dimer, indicating that this last conformation is not compatible with iron binding. All the putative hydrogen bonds drawn here are below 3.2 Å distance. c, Calculated pK_a values of conserved residues at the bottom of the crucible in the open and closed dimers. These pK_a values were calculated using PROPKA³² and the charge was deduced assuming a pH of 9.



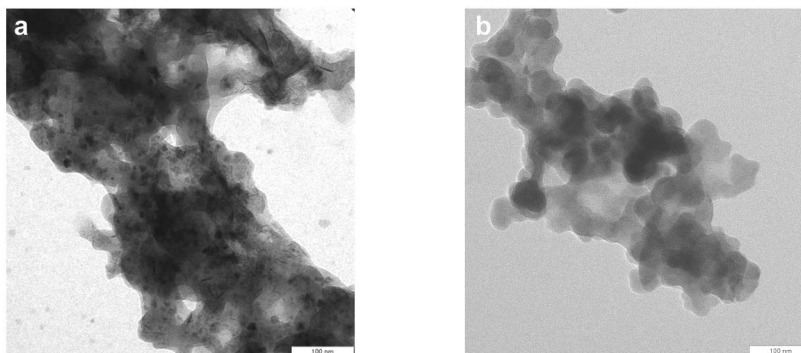
Extended Data Figure 5 | Detail of a putative hydrogen-bond network of conserved polar residues and water molecules connecting the side chain of H93 at the bottom of the crucible to the exterior of the protein. One monomer is coloured in a ramp from blue (N-Ter) to red (C-Ter), the other one is coloured in white and rendered transparent for clarity.



Extended Data Figure 6 | Size distribution of crystals determined by transmission electron microscopy. Wild type (420 particles, 28 cells), $\Delta mamP$ (425 particles, 38 cells), $\Delta mamP + mamP$ (320 particles, 29 cells), $\Delta mamP + mamP\Delta_{acid}$ (528 particles, 46 cells).

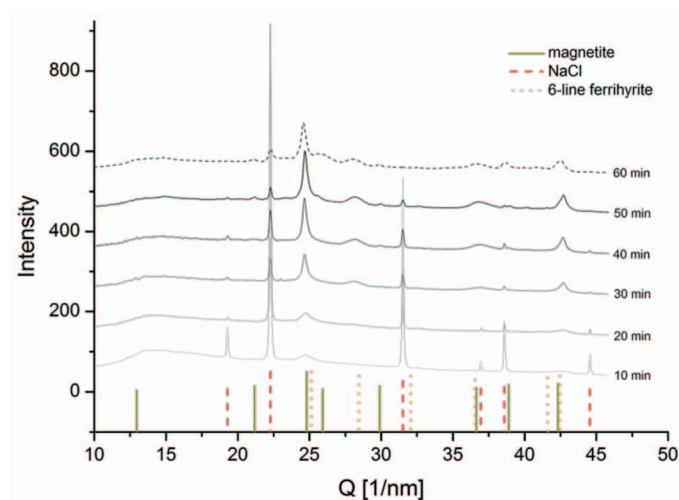


Extended Data Figure 7 | Western blot of MamP to determine expression of MamP and MamP mutant complements. The lanes are loaded as follows: whole cell extract of (1) wild type AMB-1, (2) $\Delta mamP$, (3) $\Delta mamP + mamP$, (4) $\Delta mamP + mamP_{acid}$. The antibodies were raised to a peptide of MamP of approximately 20 amino acids from strain AMB-1 (QLEGAPMILAGPRPHGYR) in rabbits by ProSci (Poway). Western blot analysis of MamP was done for each of the three biological replicates used to collect C_{mag} and TEM statistics. These images are representative of those collected for all three replicates.



Extended Data Figure 8 | TEM images indicating the presence of electron dense particles when MamP is present. **a**, Typical TEM image of the synthesis in presence of the protein. The image shows the presence electron-dense particles, probably the magnetite found by X-ray diffraction together with

poorly crystalline particulate matter. **b**, Typical TEM image of the synthesis in the absence of MamP. Only a gangue of iron ions, probably condensate from the solution while preparing the TEM grids, can be detected. These images are representative of those collected during the experiment.



Extended Data Figure 9 | Time-resolved analysis of the mineralization synthesis followed by X-ray diffraction. Reference peaks of ferrihydrite,

magnetite and sodium chloride used as salt during the synthesis and their relative intensity are indicated.

Extended Data Table 1 | Data collection, phasing and refinement statistics

Data Set	$\lambda 1$ (pk)	$\lambda 2$ (infl)	$\lambda 3$ (rem)	High Resolution	MamP Fe(II) soak
PDB Code				4JJ0	4JJ3
Data collection					
Beamlines	Proxima 1 (SOLEIL)			(SLS)	FIP (ESRF)
Space group	<i>C</i> 2			<i>C</i> 2	<i>C</i> 2
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å), β	81.1, 96.3, 54.6, 113.3°			81.2, 95.8, 54.3, 115.3°	81.4, 96.2, 54.4, 113°
Wavelength (Å)	1.7365	1.7401	1.0332	0.97627	1.7374
Resolution Range (Å)	59 – 2.5	59 – 2.5	59 – 2.5	60 – 1.8	60 – 2.8
High Resolution Range	2.64 – 2.5	2.64 – 2.5	2.64 – 2.5	1.9 – 1.8	2.95 – 2.8
Total observations	50,598	50,731	28,742	76,558	26,421
Unique reflections	13,160	13,395	11,508	30,227	9,413
Completeness (%) [*]	84.5 (80.9)	98.6 (97.9)	85.0 (85.1)	87.3 (85.9)	98.6 (99.3)
<i>I</i> / σ (<i>I</i>)	5.7 (3.1)	4.7 (1.7)	5.1 (3.0)	7.2 (2.7)	4.7 (3.0)
<i>R</i> _{sym} (%) [*]	12.2 (34.4)	14.6 (67.8)	11.0 (32.1)	8.0 (28.9)	11.9 (19.1)
<i>R</i> _{pim} (%) [†]	7.5 (20.7)	9.0 (41.8)	7.7 (21.0)	4.7 (17.3)	11.7 (18.6)
<i>R</i> _{ano} (%) [*]	8.0 (19.7)	9.5 (41.7)	7.8 (20.6)	-	10.4 (18.0)
No. of Fe sites	4	4	4	4	5
Refinement					
<i>R</i> _{cryst} / <i>R</i> _{free} (%)				18.4/24.1	26.6/29.8
No. of atoms				3,282	2,897
Protein				2,768	2,725
Water				336	-
Haem				172	172
RMSD					
Bonds (Å)				0.020	0.013
Angles (°)				2.77	2.48
Average B factors (Å ²)					
Protein				25.4	48.5
Water				28.9	-
Haem				17.9	40.6

* Values in parentheses are for the highest resolution shell.

† See ref. 13.

Single-molecule fluorescence probes dynamics of barrier crossing

Hoi Sung Chung¹ & William A. Eaton¹

Kramers developed the theory on how chemical reaction rates are influenced by the viscosity of the medium^{1,2}. At the viscosity of water, the kinetics of unimolecular reactions are described by diffusion of a Brownian particle over a free-energy barrier separating reactants and products. For reactions in solution this famous theory extended Eyring's transition state theory, and is widely applied in physics, chemistry and biology, including to reactions as complex as protein folding^{3,4}. Because the diffusion coefficient of Kramers' theory is determined by the dynamics in the sparsely populated region of the barrier top, its properties have not been directly measured for any molecular system. Here we show that the Kramers diffusion coefficient and free-energy barrier can be characterized by measuring the temperature- and viscosity-dependence of the transition path time for protein folding. The transition path is the small fraction of an equilibrium trajectory for a single molecule when the free-energy barrier separating two states is actually crossed. Its duration, the transition path time, can now be determined from photon trajectories for single protein molecules undergoing folding/unfolding transitions⁵. Our finding of a long transition path time with an unusually small solvent viscosity dependence suggests that internal friction as well as solvent friction determine the Kramers diffusion coefficient for α -helical proteins, as opposed to a breakdown of his theory, which occurs for many small-molecule reactions². It is noteworthy that the new and fundamental information concerning Kramers' theory and the dynamics of barrier crossings obtained here come from experiments on a protein rather than a much simpler chemical or physical system.

The molecule studied in this work is the all- α -helical, 73-residue, designed protein known as α_3 D (Fig. 1b and Extended Data Fig. 1). In single-molecule experiments⁶, as well as in all-atom molecular dynamics simulations⁷, α_3 D is a two-state protein (Fig. 1a) at neutral pH (Extended Data Figs 3 and 4), that is, only two states are observable at equilibrium and at all times in kinetic experiments. In the single-molecule studies reported here we use the maximum likelihood method of Gopich and Szabo⁸ in a photon-by-photon analysis of the fluorescence trajectories to obtain both the average transition path time (t_{TP}) and, from the mean residence time in the unfolded state, the folding time (t_f) (the reciprocal of the folding rate coefficient)^{6,9,10}. The basic idea of their method is to determine the most likely parameters of an assumed model that are most consistent with a collection of photon trajectories for which the colour and interval between each photon are known. For diffusive barrier crossings the times, t_f and t_{TP} , are given by the following two equations:

$$t_f = \frac{2\pi}{\beta D^* \omega^* \omega_u} \exp(\beta \Delta G_f^*) \quad (1)$$

$$t_{TP} \approx \frac{1}{\beta D^* (\omega^*)^2} \ln(2e^{\gamma} \beta \Delta G_f^*) \quad (2)$$

in which D^* is the diffusion coefficient at the free-energy barrier top, $(\omega^*)^2$ and $(\omega_u)^2$ are the curvatures of the free-energy surface at the

barrier top and the unfolded well, respectively, and ΔG_f^* is the free-energy barrier height (Fig. 1a). $\beta = 1/k_B T$, where k_B is the Boltzmann constant, T is the absolute temperature and γ is Euler's constant (0.577...). Equation (1) is from Kramers¹ and equation (2) from Szabo^{9,11}, which makes the same assumptions and approximations as Kramers concerning the underlying physics. (The major difference between Kramers and the transition state theory is that the pre-exponential factor of the latter does not contain a diffusion coefficient, and is simply $2\pi/\omega_u$ (ref. 1).) D^* ($= k_B T/\zeta^*$, the Einstein relation) is determined by the friction, ζ^* , that damps the motion across the barrier top, which in the simplest case is due entirely to solvent viscosity. Two critical assumptions in these equations are that a one-dimensional free-energy surface is sufficient to accurately describe the dynamics and that the dynamics are Brownian. That a one-dimensional free-energy surface is adequate has been validated for protein folding by lattice simulations¹², off-lattice simulations^{13,14}, and by the agreement of experiment and predictions of theoretical models^{3,4}.

An important property of equation (2) for the following development is that the transition path time is predicted to be insensitive to the

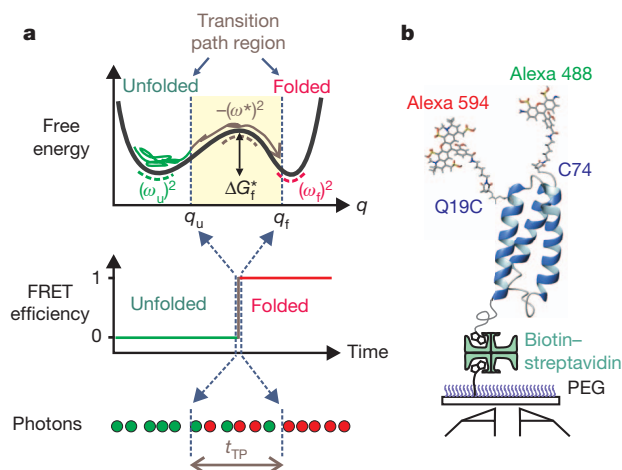


Figure 1 | Schematics of α_3 D structure and a one-dimensional free-energy surface for a two-state protein. **a**, Free energy as a function of reaction coordinate (q), a segment of a FRET efficiency trajectory, and a segment of a photon trajectory (red, acceptor; green, donor) indicating the transition path time (t_{TP}) in the idealized case in which there is no noise in the FRET efficiency in either the unfolded or folded states because they are exactly 0 and 1, respectively. In the folding reaction, the vast majority of the time is spent exploring the configurations of the unfolded well (called the waiting or residence time) with numerous unsuccessful attempts at crossing the free-energy barrier to the folded state. The diagram indicates that the 'jump' in the FRET efficiency corresponds to the transition path. The brown trajectory on a transition path is one that leaves the unfolded well, crosses the position q_u on the reaction coordinate, and reaches q_f on the other side of the barrier without re-crossing q_u . **b**, Dye-labelled α_3 D molecules are immobilized on a polyethylene glycol (PEG)-coated glass surface via a biotin-streptavidin-biotin linkage.

¹Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA.

barrier height, in contrast to the folding time, which is greatly affected by a small change of the barrier height that may be caused by viscogens and/or chemical denaturants even when the equilibrium population is unchanged¹⁵. This insensitivity of the transition path time to barrier height has very recently been observed in our single-molecule fluorescence experiments for two proteins with folding times that differ by $\sim 10^4$ -fold⁵. Consequently, the properties of the transition path time are expected to be determined by $D^*(\omega^*)^2$.

Figure 2 shows how the transition path times are determined using the Gopich–Szabo maximum likelihood method (see also Methods); Figs 3 and 4 show how the transition path and folding times depend on temperature and viscosity. One immediate result from these experiments is that the height of the free-energy barrier can be simply obtained from the ratio of the folding time to the transition path time, t_f/t_{TP} (equations (1) and (2)), if ω^*/ω_u is known. (Although energy barrier heights for reactions are routinely determined from the temperature dependence of the rate, albeit most often without consideration of the temperature-dependence of the pre-exponential factor (equation (1)), the free-energy barrier height is much more difficult to determine.) Because ω^*/ω_u cannot be obtained from our experiments, we use the value of 1.3 calculated from the potential of mean force in all-atom molecular dynamics simulations given in ref. 7 for α_3D , which together with our measurements of t_f/t_{TP} yields a ΔG_f^\ddagger of $4.2 \pm 1.0 k_B T$ at 22 °C (Fig. 3d and Methods). With a barrier height of $4.2 k_B T$, the pre-exponential factor (equation (1)) is $\sim 40 \mu s$, much larger than previous estimates of $\sim 1 \mu s$ ¹⁶. However, the larger pre-exponential factor is consistent with the longer transition path time of 10–20 μs for this protein compared with our previously measured transition path times for an all- β (2 μs) and an α/β (<10 μs) protein⁵.

For very low free-energy barriers ($\lesssim 2 k_B T$), heights have previously been estimated from ensemble kinetic experiments by Gruebele and co-workers¹⁷. In this case there is a detectable population of partially folded molecules at the barrier top. ΔG_f^\ddagger was then obtained from an approximate relation: $t_f/t_m = \exp(\beta \Delta G_f^\ddagger)$, in which t_m is a relaxation time corresponding to a ‘molecular phase’, interpreted as resulting

from a change in population at the barrier top produced by a temperature jump¹⁷. Barrier heights have also been estimated from single-molecule force experiments¹⁸, and, for very low barriers, from calorimetric measurements of the excess heat capacity¹⁹. One caveat to our measurements is that Förster resonance energy transfer (FRET) measures the transition path time for compaction of the polypeptide chain, which would underestimate the transition path time if collapse and folding are not simultaneous. For example, a twice-longer transition path time that could result from additional time for side-chain annealing within a compact structure would lower the calculated barrier, but only by $0.7 k_B T$.

As the transition path time is insensitive to the barrier height, the more interesting result is that our measurements characterize $D^*(\omega^*)^2$, the pre-logarithmic factor in equation (2). The temperature-dependent data suggest that the variation of the transition path time results from variation of D^* and not from ω^* . First, within the errors of our experiment, the small variation of the ratio t_f/t_{TP} , which does not depend on D^* , is consistent with both a constant ΔG_f^\ddagger and ω^*/ω_u over this temperature range (Fig. 3c). Furthermore, the equilibrium constant does not change with temperature (Extended Data Fig. 3 and Extended Data Table 1), suggesting that the curvatures in neither the unfolded, $(\omega_u)^2$, nor folded, $(\omega_f)^2$, wells have changed. The invariance of the free-energy surface in this temperature range is not surprising, as the protein unfolds at both lower (cold denaturation) and higher temperatures.

In the simplest case, when all of the friction (ζ^*) opposing the motion over the barrier top is due to the solvent viscosity, η , $1/D^* \propto \eta$. However, we find a much smaller viscosity dependence, that is, $1/D^* \propto \eta^\alpha$, with an α of only 0.3 (Fig. 4b). One might interpret this result as a breakdown of Kramers’ theory in which the Brownian assumption fails and causes a reduced viscosity dependence². However, given the extremely weak viscosity dependence, the more likely possibility is that there is an additional source of friction from intramolecular interactions^{20–22}. This so-called internal friction has been used previously to explain the decreased viscosity dependence observed for the relaxation rate of a protein conformational change²³ and the

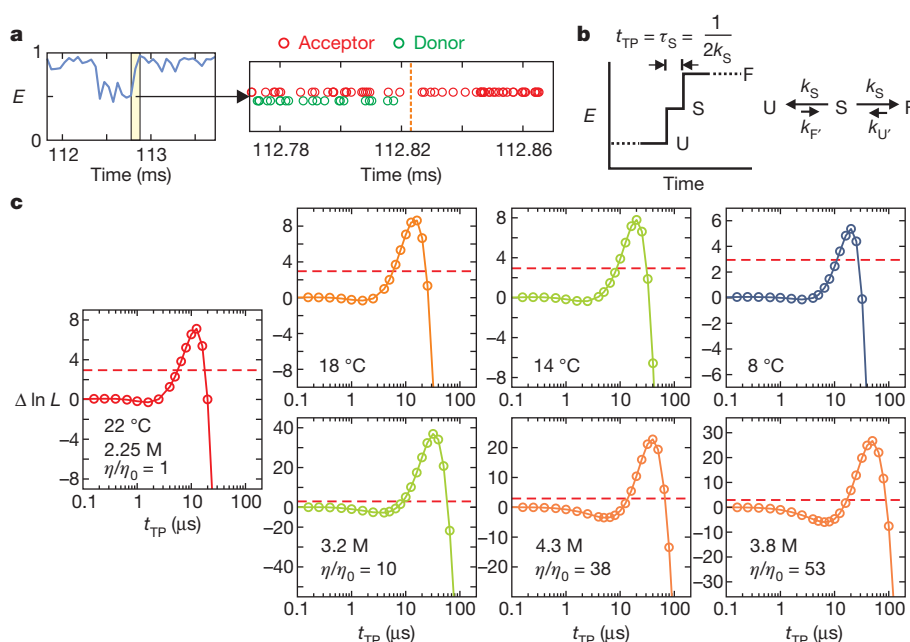


Figure 2 | Determination of transition path times. **a**, A FRET efficiency ($E = n_A/(n_A + n_D)$, where n_A is the number of acceptor photons and n_D is the number of donor photons) trajectory (50- μs bin time) with photon trajectory of the yellow segment. **b**, Schematic of a FRET efficiency trajectory using a one-step model to describe the transition path from unfolded (U) to folded (F) states for a two-state protein. The average transition path time, t_{TP} , is equal to the lifetime of a virtual intermediate state S ($\tau_S = (2k_S)^{-1}$). **c**, The difference of the log likelihood, $\Delta \ln L = \ln L(t_{TP}) - \ln L(0)$, plotted as a function of t_{TP} for folding and unfolding transitions at different temperatures (top row, 2.25 M guanidinium hydrochloride (GdmCl)) and at different solvent viscosities at 22 °C (bottom row). $L(0)$ is the likelihood for a two-state model with instantaneous folding and unfolding transitions. Therefore, $\Delta \ln L$ quantifies how much better or worse the one-step model with a finite transition path time in **b** describes the photon trajectory compared to a two-state model with an instantaneous transition (Extended Data Fig. 2b). The maximum values of $\Delta \ln L$ of all data are much greater than the 95% confidence limit (horizontal dashed line at $\Delta \ln L = 3$), which indicates that the transition path times determined by the maximum of $\Delta \ln L$ (Extended Data Tables 1 and 2) are highly statistically significant. The number of transitions analysed were 522 (22 °C), 355 (18 °C), 265 (14 °C), 284 (8 °C), 699 ($\eta/\eta_0 = 10$), 541 ($\eta/\eta_0 = 38$) and 423 ($\eta/\eta_0 = 53$).

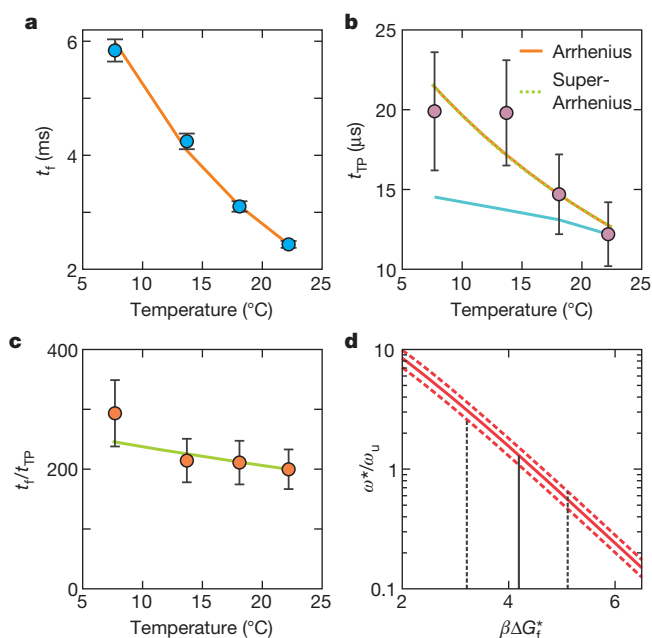


Figure 3 | Temperature dependence of folding and transition path times at 2.25 M GdmCl. **a**, Folding time. The points are the data, and the solid curve is the fit with an Arrhenius law. **b**, Transition path time. The points are the data, and the orange solid and green dashed curves are the fits with an Arrhenius and a super Arrhenius law (see text), which are indistinguishable over this small temperature range. The cyan line is the predicted transition path time scaled to the transition path time at 22 °C using equation (2) if the solvent were the only source of friction affecting the diffusion coefficient, D^* . **c**, Ratio of folding to transition path times versus temperature. The points are the data and the solid curve is the temperature dependence with $\Delta G_f^* = 4.2 k_B T$ at 22 °C obtained in **d**. **d**, Determining $\beta \Delta G_f^*$ at 22 °C. The value of ω^*/ω_u is evaluated for the measured t_f/t_{TP} , 200 (continuous red curve), including its error, that is, 233 (upper red dashed curve), and 167 (lower red dashed curve). The most probable value of ω^*/ω_u obtained from the molecular dynamics simulations⁷ is 1.3, corresponding to $\beta \Delta G_f^* = 4.2$ (see Methods).

folding time of an α -helical protein under conditions where the viscosity does not perturb the equilibrium constant, and therefore is presumed not to alter the free-energy surface¹⁵. A clear example of internal friction influencing protein dynamics can be found in the studies of Schuler and co-workers on the reconfiguration time of unfolded proteins and intrinsically disordered proteins, in which the viscosity dependence is reduced as the polypeptide becomes more compact at

the lower denaturant concentrations and increased intra-molecular interactions slow the motion of the chain²⁴.

Additional evidence for internal friction comes from the much greater temperature dependence of the transition path time than predicted by the temperature dependence of the solvent viscosity (in the absence of added viscogen) (Fig. 3b). Using the Arrhenius law, the activation energy for D^* is $11 (\pm 4) k_B T$. However, for protein folding Bryngelson and Wolynes²⁵ showed that D^* should exhibit super-Arrhenius behaviour, that is, $D^* \propto \exp[-(\Delta E/k_B T)^2]$, which fits the data equally well (Fig. 3b), in which ΔE^2 is the local mean-squared fluctuation in energy and is a measure of the underlying 'landscape roughness'. With this temperature dependence $\Delta E = 2.3 (\pm 0.4) k_B T$. Interestingly, our predicted transition path time at 370 K, the temperature of the molecular dynamics simulations⁷, is 1.7 μ s using the Arrhenius law and 2.3 μ s using the super-Arrhenius equation. Both are in excellent agreement with the observed value of 0.9 μ s in the molecular dynamics simulations. The agreement is even better if a viscosity correction were applied to the molecular dynamics value⁵, because the TIP3P water used in the simulations is threefold less viscous than real water.

Assuming that the total friction is a simple sum of the internal friction and the friction from the solvent ($\zeta = \zeta_{\text{solvent}} + \zeta_{\text{internal}}$) and that the internal friction is independent of solvent composition and viscosity^{20,23}, one might have expected that the transition path time would be linear in viscosity with a non-zero intercept (Fig. 4b), as found for the myoglobin relaxation²³ and atomistic folding simulations where the friction of the implicit solvent was varied²⁶. However, at the highest concentrations of viscogen (Extended Data Table 2), the solvent is largely non-aqueous (at $\eta/\eta_0 = 53$ the solvent is only $\sim 7\%$ water by weight), and our assumptions of solvent-independent internal friction, two-stateness and unchanged curvatures of the free-energy surface may no longer hold.

An important remaining issue is the detailed structural origin of internal friction in protein folding. What are the dominant contributions of these microscopic dynamics that give rise to internal friction and thereby slow the diffusive motion along the reaction coordinate? For dynamics involving buried residues and therefore presumably less influenced by solvent viscosity, what is the relative importance, for example, of making and breaking inter-residue contacts compared to dihedral angle flips²³? How much of the internal friction results from the increased frustration caused by local non-native contacts that is more likely for a designed protein such as α_3 D than a naturally occurring protein²⁷? Do non-Markovian effects contribute to the reduced viscosity dependence^{20,21}? Why is much larger apparent internal friction observed in folding all α -helical proteins^{15,28,29} compared to all- β or α/β proteins (see ref. 30)? The answers to these fundamental questions about protein folding dynamics will require transition path time measurements on proteins of different sequences and folds in combination with further investigations by theory and simulations.

METHODS SUMMARY

Detailed information on materials, single-molecule spectroscopy, temperature and viscosity measurements, calculation of likelihood function, and free-energy barrier height determination are described in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 May; accepted 12 September 2013.

Published online 23 October 2013.

1. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).
2. Hänggi, P., Talkner, P. & Borkovec, M. Reaction rate theory; fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).
3. Oliveberg, M. & Wolynes, P. G. The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **38**, 245–288 (2005).

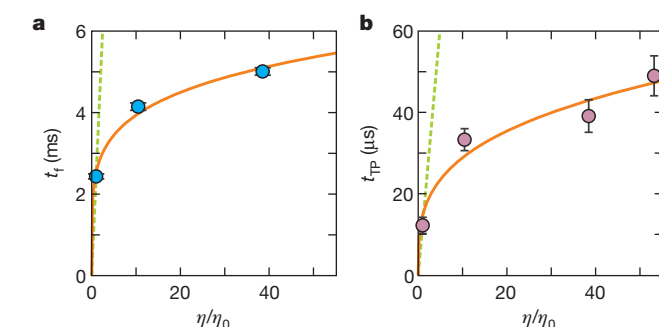


Figure 4 | Viscosity dependence of folding and transition path times. **a**, **b**, The solid curve is a fit to the power-law function $A(\eta/\eta_0)^\alpha$, with $A = 2.5 \pm 0.1$ ms and $\alpha = 0.19 \pm 0.01$ for the folding time (**a**) and $A = 15 \pm 2$ μ s and $\alpha = 0.30 \pm 0.03$ for the transition path time (**b**). The green dashed lines show the dependence expected when the folding or transition path times are linearly proportional to the solvent viscosity (see Extended Data Figs 5 and 6 and Extended Data Table 2).

4. Kubelka, J., Henry, E. R., Cellmer, T., Hofrichter, J. & Eaton, W. A. Chemical, physical, and theoretical kinetics of an ultrafast folding protein. *Proc. Natl Acad. Sci. USA* **105**, 18655–18662 (2008).
5. Chung, H. S., McHale, K., Louis, J. M. & Eaton, W. A. Single-molecule fluorescence experiments determine protein folding transition path times. *Science* **335**, 981–984 (2012).
6. Chung, H. S. *et al.* Extracting rate coefficients from single-molecule photon trajectories and FRET efficiency histograms for a fast-folding protein. *J. Phys. Chem. A* **115**, 3642–3656 (2011).
7. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
8. Gopich, I. V. & Szabo, A. Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B* **113**, 10965–10973 (2009).
9. Chung, H. S., Louis, J. M. & Eaton, W. A. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl Acad. Sci. USA* **106**, 11837–11844 (2009).
10. Chung, H. S., Cellmer, T., Louis, J. M. & Eaton, W. A. Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories. *Chem. Phys.* **422**, 229–237 (2013).
11. Hummer, G. From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **120**, 516–523 (2004).
12. Socci, N. D., Onuchic, J. N. & Wolynes, P. G. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* **104**, 5860–5868 (1996).
13. Klimov, D. K. & Thirumalai, D. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **79**, 317–320 (1997).
14. Best, R. B. & Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl Acad. Sci. USA* **102**, 6732–6737 (2005).
15. Cellmer, T., Henry, E. R., Hofrichter, J. & Eaton, W. A. Measuring internal friction of an ultrafast-folding protein. *Proc. Natl Acad. Sci. USA* **105**, 18320–18325 (2008).
16. Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding ‘speed limit’. *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).
17. Yang, W. Y. & Gruebele, M. Folding at the speed limit. *Nature* **423**, 193–197 (2003).
18. Hummer, G. & Szabo, A. Free energy surfaces from single-molecule force spectroscopy. *Acc. Chem. Res.* **38**, 504–513 (2005).
19. Godoy-Ruiz, R. *et al.* Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J. Phys. Chem. B* **112**, 5938–5949 (2008).
20. Portman, J. J., Takada, S. & Wolynes, P. G. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *J. Chem. Phys.* **114**, 5082–5096 (2001).
21. Makarov, D. E. Interplay of non-Markov and internal friction effects in the barrier crossing kinetics of biopolymers: insights from an analytically solvable model. *J. Chem. Phys.* **138**, 014102 (2013).
22. Schulz, J. C. F., Schmidt, L., Best, R. B., Dzubiella, J. & Netz, R. R. Peptide chain dynamics in light and heavy water: zooming in on internal friction. *J. Am. Chem. Soc.* **134**, 6273–6279 (2012).
23. Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J. & Eaton, W. A. The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **256**, 1796–1798 (1992).
24. Soranno, A. *et al.* Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Natl Acad. Sci. USA* **109**, 17800–17806 (2012).
25. Bryngelson, J. D. & Wolynes, P. G. Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915 (1989).
26. Zagrovic, B. & Pande, V. Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J. Comput. Chem.* **24**, 1432–1436 (2003).
27. Sutto, L., Latzer, J., Hegler, J. A., Ferreira, D. U. & Wolynes, P. G. Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc. Natl Acad. Sci. USA* **104**, 19825–19830 (2007).
28. Jas, G. S., Eaton, W. A. & Hofrichter, J. Effect of viscosity on the kinetics of α -helix and β -hairpin formation. *J. Phys. Chem. B* **105**, 261–272 (2001).
29. Wensley, B. G. *et al.* Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **463**, 685–688 (2010).
30. Hagen, S. J. Solvent viscosity and friction in protein folding dynamics. *Curr. Protein Pept. Sci.* **11**, 385–395 (2010).

Acknowledgements We are particularly indebted to J. M. Louis for the preparation, dye labelling and purification of the protein used in this work, with technical assistance from A. Aniana. We also thank R. Best, G. Hummer and A. Szabo for discussions and comments on the manuscript, and D.E. Shaw Research for providing access to their molecular dynamics trajectories for the calculations by R. Best. This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

Author Contributions H.S.C. and W.A.E. designed the research and wrote the manuscript; H.S.C. collected and analysed the experimental data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.S.C. (chunghoi@niddk.nih.gov) or W.A.E. (eaton@helix.nih.gov).

METHODS

Materials. The preparation and purification of dye-labelled α_3D has been described previously⁶.

Single-molecule spectroscopy. Single-molecule FRET experiments were performed using a confocal microscope system (MicroTime200, Picoquant). The CW mode of a dual mode (CW/pulsed) 485 nm diode laser (LDH-D-C-485, PicoQuant) was used to excite donor dyes through an oil-immersion objective (PlanApo, numerical aperture 1.4, $\times 100$, Olympus). Donor and acceptor fluorescence was collected by the same objective, split into two channels, and focused through optical filters (ET525/50m for the donor and E600LP for the acceptor, Chroma Technology) onto photon-counting avalanche photodiodes (SPCM-AQR-15, PerkinElmer Optoelectronics). Additional details for the optical setup and single-molecule experiments can be found in refs 9 and 31.

Protein molecules were immobilized on a biotin-embedded, polyethylene glycol-coated glass coverslip (Bio_01, Microsurfaces Inc.) via a biotin (surface)–streptavidin–biotin (protein) linkage. To reduce dye bleaching and blinking, 1–2 mM L-ascorbic acid (A92902, Sigma) and methyl viologen (856117, Sigma) were added to the 50 mM HEPES buffer (pH 7.6) solution³² for the temperature-dependence experiment. For the experiments in 50% and >50% glycerol, higher concentrations of 10 mM and 40 mM ascorbic acid/methyl viologen were used, respectively, to increase the bimolecular rates that reduce blinking and bleaching.

To collect a large number of trajectories, we used an automated data collection scheme as described in ref. 5.

Temperature control and measurement. Temperature was controlled by flowing dry nitrogen cooled by liquid nitrogen into a chamber surrounding the microscope objective and sample with a fixed flow rate. The temperature was varied by changing the flow rate and was measured at the confocal spot using the fluorescence lifetime of rhodamine B³³, excited by a 485-nm laser in the pulsed mode, using the calibration equation (with T in $^{\circ}C$):

$$\tau_{\text{life}} \text{ (ns)} = 2.64 - 0.0572 T + 4.63 \times 10^{-4} T^2 - 1.34 \times 10^{-6} T^3$$

Measurement of relative viscosities. The relative viscosities (η/η_0) of the solutions were obtained from the absolute temperature (T in K) and the translational diffusion times (τ) using the relationship $\eta \propto \tau T$. The translational diffusion times were measured by fluorescence correlation spectroscopy experiments on an Alexa Fluor 488- and Alexa Fluor 594-labelled rigid rod molecule, 20-residue poly-proline, as described previously⁵.

Calculation of donor and acceptor cross-correlation. To test further the assumption of a two-state model in our data analysis using the maximum likelihood method (see below), we calculated the donor–acceptor cross-correlation function, as $C_{DA}(\tau) = -A \exp(k\tau)$, as shown in Extended Data Fig. 4 and compared its decay rate with the sum of the rate coefficients determined by the maximum likelihood method (Extended Data Table 1). To obtain the timescale of the unfolded state dynamics at high viscosity, the donor–acceptor cross-correlation function was calculated (Extended Data Fig. 6) for the photon trajectories in the unfolded segments that were extracted using the Viterbi algorithm^{6,34,35}.

Calculation of likelihood functions. To determine parameters for the two-state kinetics, we used the Gopich–Szabo maximum likelihood method and analysed the photon trajectories without time binning. This method yields the most likely parameters of an assumed model that can reproduce the observed photon trajectories. The likelihood function for the j th photon trajectory is⁸:

$$L_j = I^T \prod_{i=2}^{N_j} [F(c_i) \exp(K\tau_i)] F(c_1) \mathbf{p}_{\text{eq}} \quad (3)$$

in which N_j is the number of photons in a trajectory, c_i is the colour of the i th photon (donor or acceptor) and τ_i is a time interval between the i th and $(i-1)$ th photons (Extended Data Fig. 2a). The photon colour matrix F depends on the colour of a photon as $F(\text{acceptor}) = E$ and $F(\text{donor}) = I - E$, where E is a diagonal matrix with elements that are FRET efficiencies of the individual states. I is the identity matrix, \mathbf{p}_{eq} is a vector consisting of the equilibrium population of each state, and I^T is a row vector with elements of 1. The calculation of likelihood values was performed using the diagonalization of the matrix exponential in equation (3) as described in ref. 8. Practically, the log likelihood function was calculated and the total log likelihood function of all trajectories was calculated by summing individual log likelihood functions as $\ln L = \sum_j \ln L_j$. For a two-state system, there are

four independent parameters—the apparent FRET efficiencies and the rate coefficients of the folded and unfolded states (E_F , E_U , k_F , k_U). In analysing the data, instead of varying the two rate coefficients, it is more convenient to vary the sum of the rate coefficients ($k = k_F + k_U$) and the fractional population of folded molecules ($p_F = k_F/(k_F + k_U)$). For the two-state kinetic model (Extended Data Fig. 2b), the rate matrix is given by:

$$K = \begin{pmatrix} -k_U & k_F \\ k_U & -k_F \end{pmatrix} \quad (4)$$

The four parameters (E_F , E_U , k_F , k_U) can be obtained by maximizing the log likelihood $\ln L$.

In the measurement of the average transition path time, t_{TP} , molecules were illuminated at 10–15 times higher laser intensity ($20\text{--}30 \text{ kJ cm}^{-2}$) than the intensity used for the determination of two-state kinetics parameters above to collect photons at high photon count rates of $550\text{--}710 \text{ ms}^{-1}$. This high photon count rate permitted clear resolution of transitions by collecting photons in 50–100- μs bins. The photon trajectory in the short segment near the transition ($\sim 1 \text{ ms}$) was then analysed using the likelihood function. We adopted a three state model with a virtual intermediate ('step') state S in addition to the folded (F) and unfolded (U) states, which is the simplest discrete representation of how the FRET efficiency changes along the transition path. In this model, the FRET efficiency of S is midway between the folded and unfolded states ($E_S = (E_F + E_U)/2$) and the lifetime of S (τ_S) corresponds to the average transition path time, t_{TP} . The rate matrix of this kinetics scheme (Extended Data Fig. 2c) is:

$$K = \begin{pmatrix} -k_{U'} & k_S & 0 \\ k_{U'} & -2k_S & k_{F'} \\ 0 & k_S & -k_{F'} \end{pmatrix} \quad (5)$$

F' and U' are used for the folded and unfolded states to distinguish these from those in the two-state model with an instantaneous transition. In this analysis $t_{TP} = \tau_S = 1/2k_S$, and t_{TP} is the same for folding and unfolding transitions, which is also the case for the real transition path in Fig. 1a, regardless of the relative stability of the folded and unfolded states as long as q_u and q_f are the same for the two paths. The likelihood function of this model for the j th photon trajectory with a single transition is⁸:

$$L_j = \mathbf{v}_{\text{fin}}^T \prod_{i=2}^{N_j} [F(c_i) \exp(K\tau_i)] F(c_1) \mathbf{v}_{\text{ini}} \quad (6)$$

in which \mathbf{v}_{ini} and \mathbf{v}_{fin} are state vectors at the beginning (ini) and the end (fin) of the trajectory. For the case of a folding transition in Extended Data Fig. 2a, $\mathbf{v}_{\text{ini}} = (0 \ 0 \ 1)^T$ and $\mathbf{v}_{\text{fin}} = (1 \ 0 \ 0)^T$. In the analysis, the folded and unfolded segments were assigned using the Viterbi algorithm^{34,35}, adapted for photon trajectories⁶, and k_F and k_U were obtained from the low-intensity measurements (Extended Data Table 1). When there are multiple transitions in a trajectory, each pair of segments with a single folding or unfolding transition was analysed separately. In the calculation of the likelihood function in equation (6) for these segments, we reduced the rate coefficients by a factor of 1,000 to effectively eliminate the possibility of multiple transitions that are not resolvable, that is, $k_{U'}$ ($= k_U/1,000$) and $k_{F'}$ ($= k_F/1,000$). This treatment is valid because we use the difference of the log likelihood values ($\Delta \ln L$) in this paper.

The reader is referred to previous publications for an in-depth understanding of the Gopich–Szabo maximum likelihood method⁸ and its experimental application to a two-state protein system⁶ and for the transition path analysis⁵.

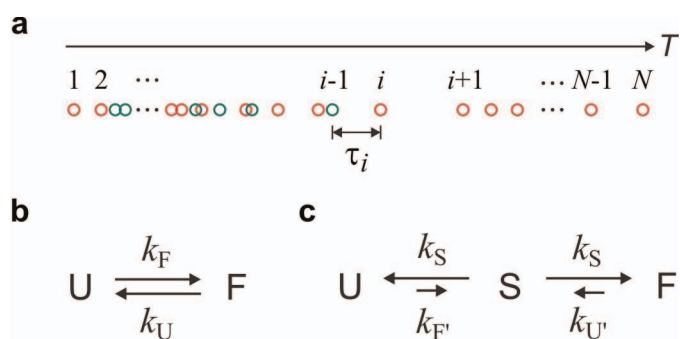
Determination of the free energy barrier height. The free-energy barrier height at $22^{\circ}C$ was determined (in Fig. 3d) using the ratio of the folding time to the transition path time t_f/t_{TP} , and the ratio of the curvatures at the barrier top and the bottom of the unfolded state (ω^*/ω_u)². The value of ω^*/ω_u was evaluated for the measured t_f/t_{TP} , 200 (continuous red curve), including its error, that is, 233 (top red dashed curve), and 167 (bottom red dashed curve). The most probable value of ω^*/ω_u obtained from the molecular dynamics simulations⁷ is 1.3, the average of 0.94 (using the fraction of native contacts as the reaction coordinate³⁶ and 1.6 (using an optimized reaction coordinate⁷), corresponding to $\beta\Delta G_f^{\ddagger} = 4.2$. Assuming an uncertainty of a factor of 2 for ω^*/ω_u , the possible range of $\beta\Delta G_f^{\ddagger}$ becomes 3.2 to 5.1 for ω^*/ω_u between 0.65 and 2.6. Using the reported times from the molecular dynamics simulations of $t_f = 27 \pm 8 \mu\text{s}$, and $t_{TP} = 0.9 \pm 0.2 \mu\text{s}$, and $\omega^*/\omega_u = 1.6$ from the one-dimensional free-energy surface constructed with an optimized reaction coordinate, the ratio of equations (1) and (2) yields a barrier height of $1.7 \pm 0.6 k_B T$, which is close to the $2.1 k_B T$ (no error reported) barrier height for this surface⁷. Using the fraction of native contacts as the reaction coordinate for all three quantities, the corresponding times are $t_f = 26.4 \pm 8.8 \mu\text{s}$, $t_{TP} = 0.83 \pm 0.17 \mu\text{s}$ (boundaries for the transition path are the two well minima), and $\omega^*/\omega_u = 0.94 \pm 0.18$, the calculated barrier height is $2.5 \pm 0.6 k_B T$, compared to the barrier height of $2.4 \pm 0.3 k_B T$ (ref. 36).

31. Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V. & Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl Acad. Sci. USA* **104**, 1528–1533 (2007).
32. Vogelsang, J. *et al.* A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem.* **47**, 5465–5469 (2008).

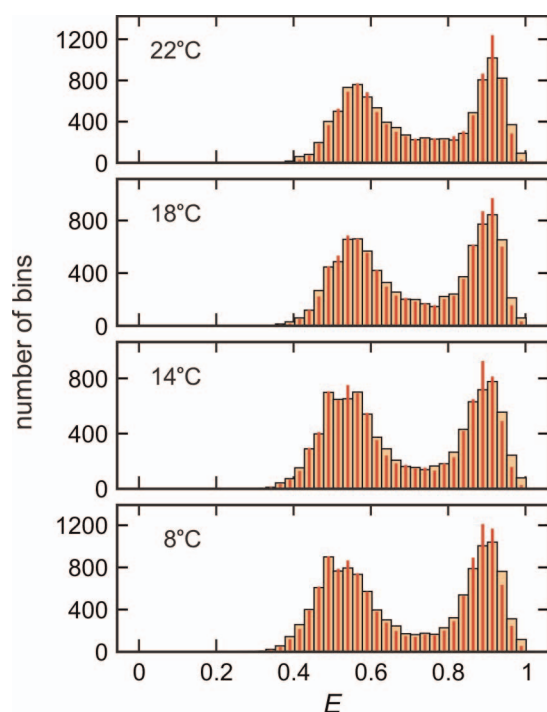
33. Benninger, R. K. P. *et al.* Quantitative 3D mapping of fluidic temperatures within microchannel networks using fluorescence lifetime imaging. *Anal. Chem.* **78**, 2272–2278 (2006).
34. Viterbi, A. J. Error bounds for convolution codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory* **13**, 260–269 (1967).
35. Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech. *Proc. IEEE* **77**, 257–286 (1989).
36. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl Acad. Sci. USA*. <http://dx.doi.org/10.1073/pnas.1311599110> (in the press).
37. Zhu, Y. *et al.* Ultrafast folding of α_3 D: a *de novo* designed three-helix bundle protein. *Proc. Natl Acad. Sci. USA* **100**, 15486–15491 (2003).
38. Liu, F. *et al.* A one-dimensional free energy surface does not account for two-probe folding kinetics of protein α_3 D. *J. Chem. Phys.* **130**, 061101 (2009).

AviTag	Linker
GMSG LNDIFE AQ	KIEWHESS GLVPRGSHM
MGSWAEFKQR LAAIKTRL	CA LGGSEAELAA FEKEIAAFES
ELQAYKGKGN PEVEALRKEA	AAIRDELQAY RHNC α_3 D

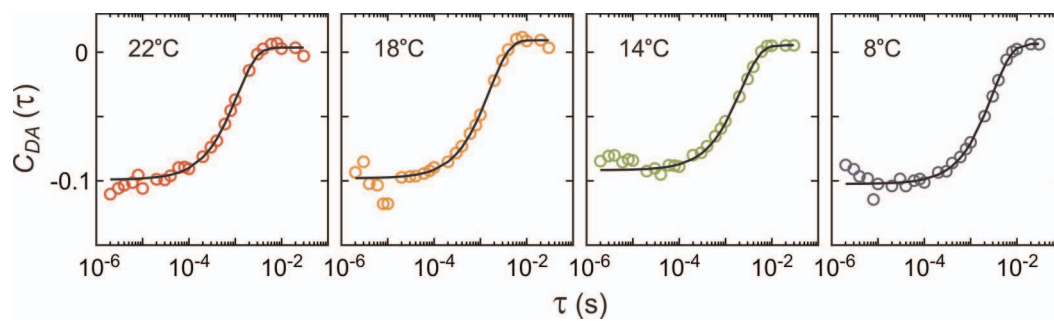
Extended Data Figure 1 | Amino acid sequences of polypeptides containing protein α_3 D. Dyes were attached to the cysteine residues (red) and a biotin molecule was attached to the lysine residue (blue) in the AviTag sequence.



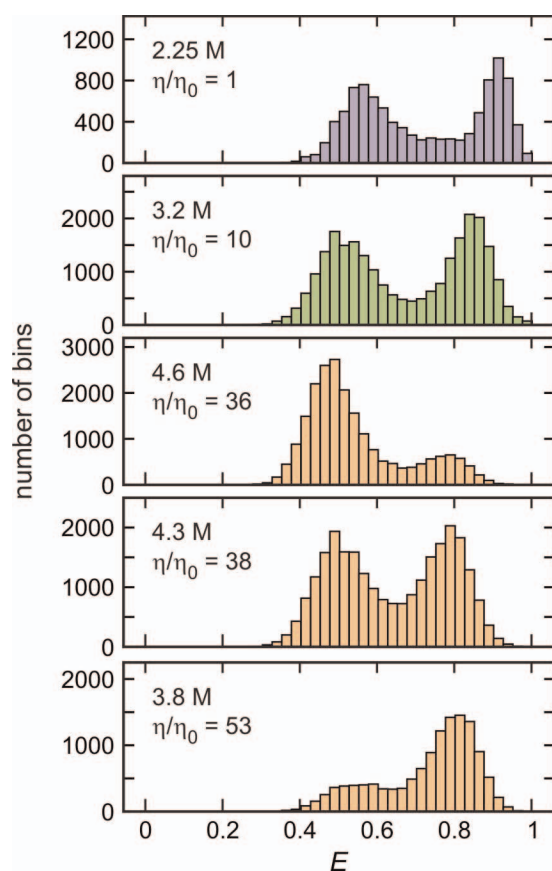
Extended Data Figure 2 | Photon trajectory and kinetics models. **a**, The definition of photon indices and time interval of a photon trajectory with a folding transition. **b**, **c**, Photon trajectories were analysed using the two-state model to determine kinetics parameters (**b**) or the three-state model to determine the average transition path times ($t_{TP} = 1/2k_S$) (**c**).



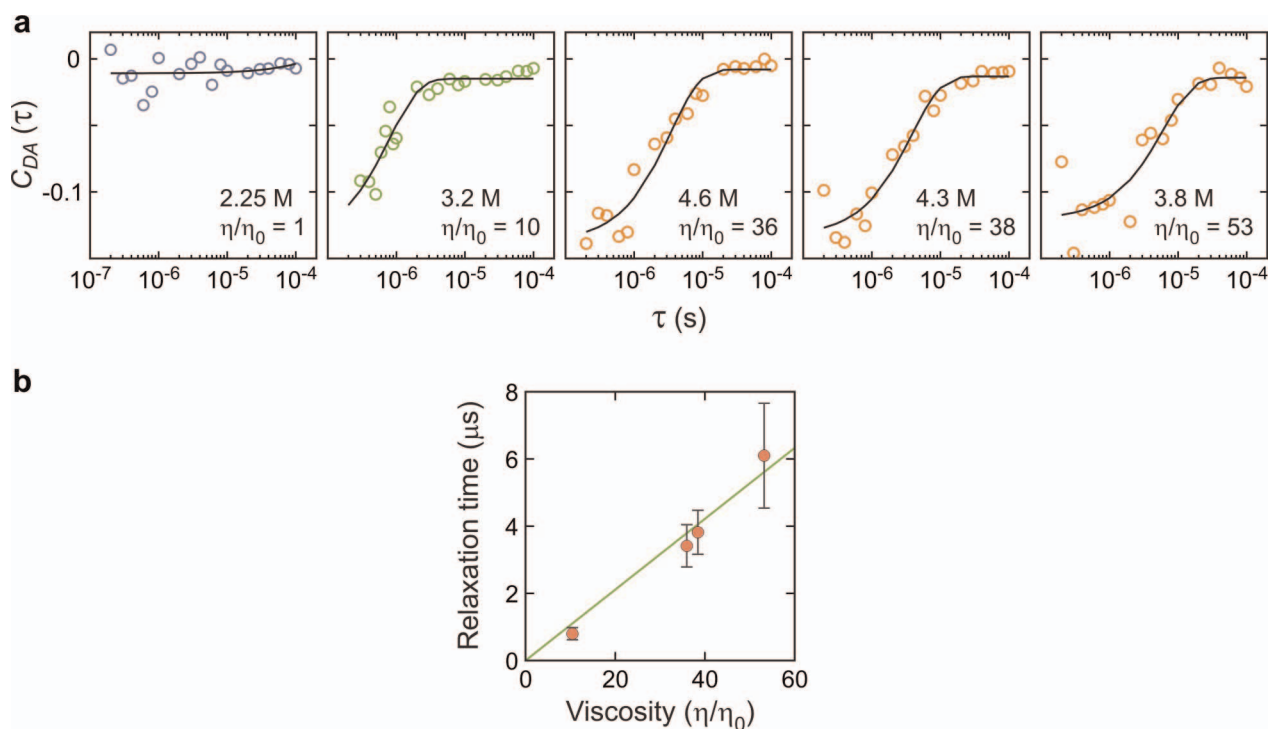
Extended Data Figure 3 | FRET efficiency histograms of α_3 D in 2.25 M GdmCl solution at different temperatures. The FRET efficiency histograms were constructed from 1-ms bins in the trajectories with the mean photon count rate $>40 \text{ ms}^{-1}$. Wide and narrow bars are the experimental histograms and the histograms constructed from re-coloured photon trajectories using the parameters obtained from the maximum likelihood method with the two-state model (Extended Data Table 1), respectively. The agreement between the two histograms validates the description of α_3 D as a two-state folder⁸. The similar ratio of the integral of the folded (high FRET) and the unfolded (low FRET) distributions indicates that the equilibrium constant is unchanged over the temperature range of the measurement, as shown more precisely in the maximum likelihood analysis. At high temperature and low pH, where the 11 glutamates and 1 aspartate are protonated, more than two states are observed^{37,38}.



Extended Data Figure 4 | Donor-acceptor cross-correlation functions at different temperatures. Black solid lines are exponential functions that best fit the data. The fitting parameters are listed in Extended Data Table 1.



Extended Data Figure 5 | FRET efficiency histograms of α_3D at various solvent viscosities. The FRET efficiency histograms were constructed from 1-ms bins in the trajectories with the mean photon count rate $>50 \text{ ms}^{-1}$ for 2.25 M and 3.2 M GdmCl and from 2-ms bins in the trajectories with the mean photon count rate $>30 \text{ ms}^{-1}$ for 4.6 M, 4.3 M and 3.8 M GdmCl concentrations. At the relative viscosity (η/η_0) 1, 10 and 38, the higher concentration of GdmCl was used to counteract the stabilization of proteins by glycerol to maintain the ratio of folded to unfolded molecules as close to unity as practically possible. The similar ratio of the integral of the folded (high FRET) and the unfolded (low FRET) distributions indicates that the equilibrium constant is unchanged at these conditions, as shown more precisely in the maximum likelihood analysis.



Extended Data Figure 6 | Donor-acceptor cross-correlation of the segments of the fluorescence trajectories corresponding to the unfolded state²⁴. **a**, Black solid lines are exponential functions that best fit the data. The fitting parameters are listed in Extended Data Table 2. **b**, The unfolded state

dynamics are slowed approximately linearly by the solvent viscosity as previously observed at high denaturant concentrations²⁴. The relaxation time at $\eta/\eta_0 = 1$ (aqueous solution) is too fast to be measured by this method.

Extended Data Table 1 | Temperature dependence of the kinetic parameters obtained from the two-state maximum likelihood analysis, the relaxation rate obtained from the donor-acceptor cross-correlation analysis, and the transition path times

Temperature (°C)	22.2	18.1	13.7	7.7
Viscosity (η/η_0)	1	1.07	1.12	1.19
E_F	0.91 (0.0005)	0.90 (0.0006)	0.89 (0.0006)	0.89 (0.0006)
E_U	0.55 (0.0007)	0.54 (0.0008)	0.52 (0.0008)	0.52 (0.0008)
k (ms ⁻¹)	0.85 (0.02)	0.66 (0.02)	0.52 (0.01)	0.37 (0.01)
ρ_F	0.48 (0.008)	0.49 (0.009)	0.46 (0.010)	0.47 (0.010)
Donor-acceptor cross correlation, k (ms ⁻¹)	0.92 (0.04)	0.64 (0.02)	0.47 (0.02)	0.34 (0.01)
Transition path time (μ s)	12.2 (2.0)	14.7 (2.5)	19.8 (3.3)	19.9 (3.7)

Errors are standard deviations obtained from the diagonal elements of the covariance matrix calculated from the likelihood function. [GdmCl] = 2.25 M.

Extended Data Table 2 | Viscosity dependence of the kinetic parameters obtained from the two-state maximum likelihood analysis, the transition path time and the correlation time of the unfolded state dynamics obtained from the donor–acceptor cross-correlation at 22 °C. Errors are standard deviations obtained from the diagonal elements of the covariance matrix calculated from the likelihood function

GdmCl (M)	2.25	3.2	4.6	4.3	3.8
Glycerol (%)	0	50	55	58	61
Viscosity (η/η_0)	1	10	36	38	53
E_F	0.91 (0.0005)	0.84 (0.0004)	0.79 (0.0007)	0.79 (0.0004)	0.81 (0.0005)
E_U	0.55 (0.0007)	0.50 (0.0005)	0.47 (0.0004)	0.49 (0.0005)	0.52 (0.0011)
k (ms^{-1})	0.85 (0.02)	0.48 (0.01)	0.46 (0.01)	0.40 (0.01)	0.56 (0.01)
p_F	0.48 (0.008)	0.50 (0.006)	0.23 (0.004)	0.50 (0.005)	0.72 (0.005)
Transition path time (μs)	12.2 (2.0)	33.3 (2.7)	NA	39.1 (4.0)	49.0 (4.9)
Donor-acceptor cross correlation time in the unfolded state, $1/k$ (μs)	NA	0.80 (0.18)	3.4 (0.6)	3.8 (0.7)	6.1 (1.6)

Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock

Michael J. Haydon^{1†}, Olga Mielczarek¹, Fiona C. Robertson^{1†}, Katharine E. Hubbard¹ & Alex A. R. Webb¹

Circadian clocks provide a competitive advantage in an environment that is heavily influenced by the rotation of the Earth^{1,2}, by driving daily rhythms in behaviour, physiology and metabolism in bacteria, fungi, plants and animals^{3,4}. Circadian clocks comprise transcription–translation feedback loops, which are entrained by environmental signals such as light and temperature to adjust the phase of rhythms to match the local environment³. The production of sugars by photosynthesis is a key metabolic output of the circadian clock in plants^{2,5}. Here we show that these rhythmic, endogenous sugar signals can entrain circadian rhythms in *Arabidopsis thaliana* by regulating the gene expression of circadian clock components early in the photoperiod, thus defining a ‘metabolic dawn’. By inhibiting photosynthesis, we demonstrate that endogenous oscillations in sugar levels provide metabolic feedback to the circadian oscillator through the morning-expressed gene *PSEUDO-RESPONSE REGULATOR 7* (*PRR7*), and we identify that *prp7* mutants are insensitive to the effects of sucrose on the circadian period. Thus, photosynthesis has a marked effect on the entrainment and maintenance of robust circadian rhythms in *A. thaliana*, demonstrating that metabolism has a crucial role in regulation of the circadian clock.

In plants, energy is derived from photosynthesis in chloroplasts by fixing CO₂ into sugar in a light-dependent manner. Net carbon assimilation and starch metabolism are under circadian regulation^{2,6–8}, as are transcripts associated with chlorophyll biosynthesis and the photosynthetic apparatus, which peak at about 4 h after dawn⁵. For *A. thaliana* seedlings, the addition of sucrose to the growth medium shortens the circadian period when the plants are grown in continuous light⁹ and can sustain circadian rhythms when they are grown in continuous dark¹⁰. Because exogenous sugars can influence the circadian oscillator, we sought to investigate whether endogenous sugars derived from photosynthesis are part of the circadian network in plants.

To investigate whether photosynthesis can influence the core circadian clock in *A. thaliana*, we inhibited photosynthesis by growing seedlings in CO₂-free air or in medium containing 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU), an inhibitor of photosystem II, and we monitored the circadian rhythms of luciferase (LUC)-based reporters of transcription of the core clock gene promoters. We performed these experiments in continuous low light (10 μmol m⁻² s⁻¹) because we observed that exogenous sucrose markedly shortened the circadian period in low light but had only subtle effects in higher light (50 μmol m⁻² s⁻¹) (Extended Data Fig. 2a, b). CO₂ depletion (Fig. 1a) or DCMU treatment (Extended Data Fig. 2c, d) increased the period of the clock reporters by a mean length of 2.9 h and 2.5 h, respectively, compared with untreated controls. Both treatments increased the activity of the *PRR7* promoter: *LUC* construct (*PRR7:LUC*) and reduced the activity of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) promoter: *LUC* construct (*CCA1:LUC*), which then damped towards arrhythmia over time (Fig. 1a and Extended Data Fig. 2b). In contrast to exogenous sucrose, the effects of DCMU were similar in high light and low light (Fig. 1b and Extended Data Fig. 2b). Exogenous sucrose is probably

ineffective at altering the period in higher light because the response is already saturated by higher levels of endogenous sugars produced by photosynthesis, whereas complete inhibition of photosynthesis will be effective at altering the period under either set of light conditions.

In light–dark cycles, there are robust endogenous rhythms of soluble sugars, which peak at about 4–8 h after dawn^{6,11} (Extended Data Fig. 3a). Inhibition of photosynthesis by either CO₂ depletion or DCMU treatment reduced the endogenous glucose, fructose and sucrose concentrations (Extended Data Fig. 3b, c). To test whether the effects of inhibition of photosynthesis on the circadian oscillator were due to reduced sugar production, we resupplied exogenous sucrose to CO₂-depleted or DCMU-treated seedlings. The period lengthening normally induced

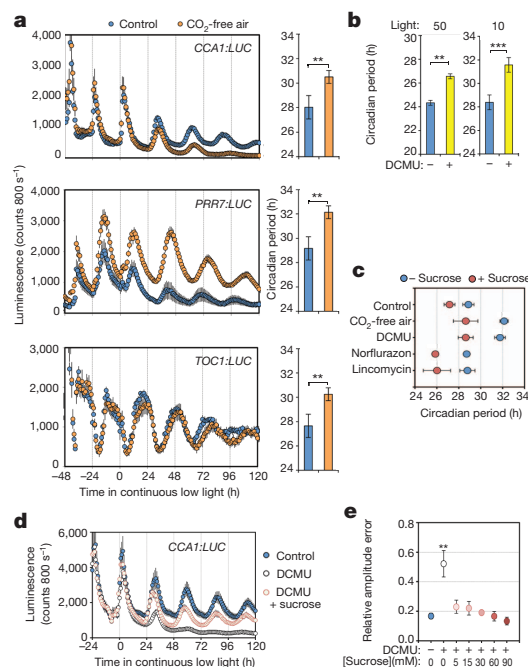


Figure 1 | Photosynthetically derived sugars influence the circadian clock in *A. thaliana*. **a**, LUC reporter activity rhythms (mean \pm s.e.m.) (left) and period estimates (right) for seedlings grown in ambient air (control) or CO₂-free air in continuous low light (10 μmol m⁻² s⁻¹) ($n = 4$). **b**, Period estimates for *PRR7:LUC* activity rhythms in continuous light (10 μmol m⁻² s⁻¹ or 50 μmol m⁻² s⁻¹) in the presence or absence of DCMU (mean \pm s.d.; $n = 4$). **c**, Period estimates for *PRR7:LUC* activity rhythms for seedlings grown in continuous low light and treated with inhibitors in the presence or absence of exogenous sucrose (mean \pm s.d.; $n = 4$). **d**, **e**, *CCA1:LUC* activity rhythms (mean \pm s.e.m.) (**d**) and relative amplitude error of *CCA1:LUC* activity rhythms (**e**) in seedlings treated with DCMU in the presence or absence of exogenous sucrose (mean \pm s.d.; $n = 4$). ** $P < 0.01$; *** $P < 0.001$ compared with untreated controls; two-tailed Student's *t*-test. n refers to number of biological replicates.

¹Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. †Present addresses: Department of Biology, University of York, York YO10 5DD, UK (M.J.H.); Department of Biochemistry, University of Zimbabwe, PO Box MP45, Harare, Zimbabwe (F.C.R.).

by either treatment was suppressed by the addition of exogenous sucrose (Fig. 1c). The effects of DCMU treatment on *CCA1:LUC* rhythms were reversed by the addition of as little as 5 mM (0.15% w/v) exogenous sucrose to the growth medium (Fig. 1d, e and Extended Data Fig. 4a). We also tested the effect of norflurazon or lincomycin, both of which trigger retrograde signalling from the chloroplast to the nucleus¹². Neither treatment lengthened the circadian period of activity of *PRR7:LUC* (Fig. 1c) or inhibited the activity of *CCA1:LUC* (Extended Data Fig. 4b) in the presence or absence of exogenous sucrose. Furthermore, we found no evidence that photosynthesis might affect clock function through mechanisms associated with reactive oxygen species production (Extended Data Fig. 5), which agrees with a recent report¹³.

Because our data suggest that the effects of photosynthesis on the circadian clock are mediated by sugars, we investigated the role of sugars in circadian function in more detail. We first tested whether the effects of exogenous sucrose represent a general response to sugar. In continuous low light, the circadian periods of *CCA1:LUC*, *PRR7:LUC* and *TIMING OF CAB EXPRESSION 1 (TOC1)* promoter:*LUC* activity were a mean of 4.2 h shorter in seedlings grown in medium containing 90 mM (3% w/v) sucrose, glucose or fructose than in mannitol-treated controls (Fig. 2a, b). Similarly, exogenous sucrose, glucose or fructose,

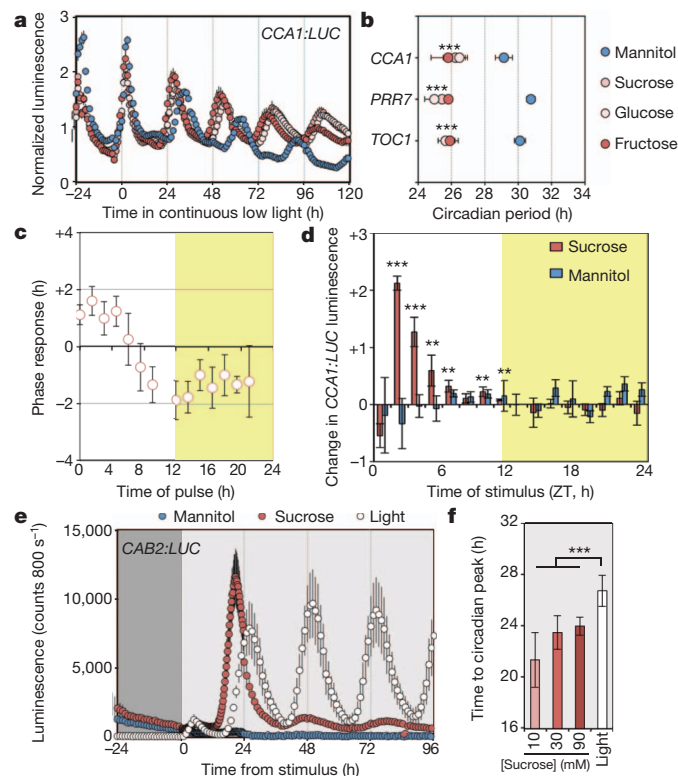


Figure 2 | Metabolically active sugar is a zeitgeber that acts differently from light. **a, b,** *CCA1:LUC* activity rhythms (mean \pm s.e.m.) (**a**) and period estimates (**b**) in seedlings grown in continuous low light with the indicated sugar ($n = 4$). **c,** The phase response of *CCA1:LUC* activity rhythms to pulses of sucrose in continuous low light (mean \pm s.d.; $n = 8$). **d,** The change in normalized *CCA1:LUC* activity 3 h after sucrose or mannitol treatment in continuous low light (mean \pm s.d.; $n = 4$). Shading in **c** and **d** represents subjective night. **e,** *CAB2:LUC* activity rhythms in seedlings dark-adapted for 72 h, then treated with sucrose or mannitol in continuous dark or transferred to continuous light (mean \pm s.e.m.; $n = 4$). Dark grey represents constant dark, and light grey represents either constant dark plus sucrose treatment or constant light. **f,** Time to the first circadian peak of *CAB2:LUC* activity in seedlings from **e**. After treatment with sucrose at various concentrations or light (mean \pm s.d.; $n = 4$). ** $P < 0.01$; *** $P < 0.001$ compared with mannitol-treated plants (in **b** and **d**); two tailed Student's *t*-test. n refers to number of biological replicates.

but not mannitol or a non-metabolizable glucose analogue (3-*O*-methyl glucose), were able to sustain circadian rhythms in continuous dark (Extended Data Fig. 6a). These data suggest that the effects of exogenous sucrose on circadian rhythms represent a general response to metabolically active sugars.

Oscillations of circadian reporter activity are absent or very low in continuous dark¹⁰ (Extended Data Fig. 6b). Exogenous sucrose can reinitiate the circadian oscillations of the clock output reporter *CHOROPHYLL A/B BINDING PROTEIN 2 (CAB2)* promoter:*LUC* in dark-adapted seedlings, and the phase is set to the time of sucrose addition after 72 h (subjective dawn) or 60 h (subjective dusk) in continuous dark¹⁰. We observed the same behaviour for reporters of the core circadian oscillator and confirmed that exogenous sucrose led to an increased *CCA1* transcript level in dark-adapted seedlings (Extended Data Fig. 7). The phase setting of the clock indicates that sucrose is not simply amplifying damped rhythms in dark-adapted seedlings as a result of the increased availability of ATP, and this finding suggests a role for sugars in entrainment. To directly test whether sugars function in entrainment, we determined a phase-response curve (PRC) for exogenous sucrose; this curve assesses the ability of a stimulus to alter the circadian phase across a circadian cycle¹⁴. In continuous low light, the phase of the peak activity of *CCA1:LUC* and *TOC1:LUC* was shifted by pulses of exogenous sucrose, inducing phase advances of up to 2 h around dawn and phase delays around dusk (Fig. 2c and Extended Data Fig. 8). We observed subtle differences between reporters, similar to phase setting by temperature¹⁵. The phase shifts were not due to effects on circadian period or an osmotic signal (Extended Data Fig. 8c). These data are consistent with metabolically active sugars functioning as a Type 1 zeitgeber participating in circadian entrainment¹⁴.

A key feature of entrainment is variation, or 'gating', of the response to the zeitgeber in a time-dependent manner³. Sucrose application during the first subjective day of continuous low light significantly induced *CCA1:LUC* activity during the day but had little effect during the subjective night (Fig. 2d). This effect was most pronounced before midday (zeitgeber time 6 (ZT6)). These data demonstrate that the input of sugars to *CCA1:LUC* activity is gated so that the promoter is most responsive to sucrose availability early in the light period.

We next compared the responses of the *A. thaliana* circadian system to sucrose and light because light can act as a strong, Type 0 zeitgeber^{3,14} and drives sugar production from photosynthesis. In dark-adapted seedlings, there was a similar transient increase in *CAB2:LUC* activity, peaking at about 5 h after treatment with light or sucrose (Fig. 2e). By contrast, the first circadian peak in *CAB2:LUC* occurred 26.9 h after the onset of light, compared with 22.8 h after sucrose addition, indicating a 4.1 h advanced phase set by sucrose compared with light (Fig. 2e, f). The difference in phase setting did not depend on sucrose concentration or light intensity within the ranges tested (Fig. 2f and Extended Data Fig. 8d). When photosynthetic sugar production was inhibited in the light by DCMU, the phase that was set by light was delayed by a further 2.5–3.5 h (Extended Data Fig. 8e). These data demonstrate that these zeitgebers both function in discrete (non-parametric) entrainment. The difference in phase might be due to period effects but could also indicate distinct phase setting. The phase difference coincides with the delay between dawn and the highest endogenous sucrose and glucose concentrations (Extended Data Fig. 3). We propose that a concentration threshold of photosynthetically derived sugars provides input to the central oscillator, thereby functioning as a 'metabolic dawn' that contributes to entrainment of the *A. thaliana* circadian clock (Extended Data Fig. 1).

Having established that sugars derived from photosynthesis contribute to circadian entrainment in *A. thaliana*, we next investigated how this might occur. The increase in *PRR7:LUC* activity in DCMU-treated or CO₂-depleted seedlings (Fig. 1a and Extended Data Fig. 2) suggested that photosynthesis regulates *PRR7* transcript abundance. We measured the transcript levels of morning-expressed circadian clock genes in the shoots of control (untreated) and DCMU-treated

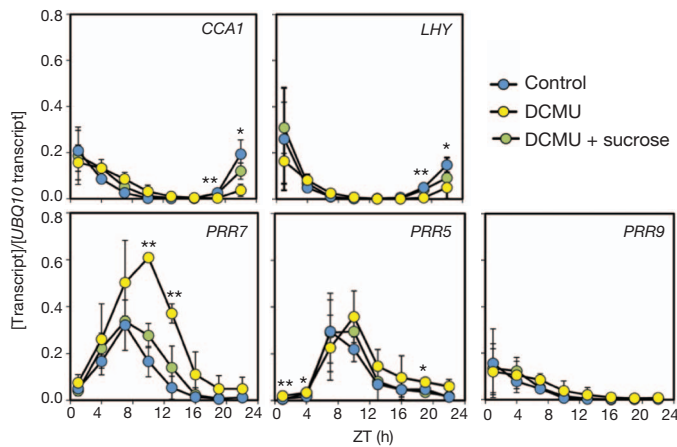


Figure 3 | Photosynthetically derived sugar represses *PRR7* late in the photoperiod. Transcript levels of circadian clock genes, relative to *UBQ10* expression, in the leaves of seedlings treated with DCMU in the presence or absence of exogenous sucrose, under 12 h light–12 h dark conditions (mean \pm s.d.; $n = 3$). * $P < 0.05$; ** $P < 0.01$; difference between DCMU-treated and untreated (control) plants; two-tailed Student's *t*-test. *n* refers to number of biological replicates.

seedlings (Fig. 3). *PRR7* transcript levels were 3.7-fold to 8.2-fold higher in DCMU-treated seedlings than in untreated plants between ZT10 and ZT16, and this difference was suppressed when sucrose was added to the medium. *PRR5* transcript levels were only 1.6-fold to 2.9-fold higher in DCMU-treated seedlings around dawn (ZT0) than in untreated plants, and *PRR9* transcript levels were unaffected. *CCA1* and *LATE ELONGATED HYPOCOTYL (LHY)* transcripts were 3.0-fold to 8.4-fold lower before dawn, between ZT18 and ZT21, in DCMU-treated seedlings than in untreated material. These data are consistent with the *LUC* reporter data (Extended Data Fig. 2) and suggest that the effect of photosynthesis is more pronounced on *PRR7* than on the other tested circadian clock genes.

These data led us to propose that photosynthetic input to the circadian oscillator might act through *PRR7*, a transcriptional repressor that acts on the *CCA1* promoter during the night¹⁶. We first tested the short-term effect of exogenous sucrose on *PRR7* promoter activity. In contrast to *CCA1:LUC* (Fig. 2d), *PRR7:LUC* activity was significantly repressed during the day and the subjective night, but this repression was most pronounced during the morning (Extended Data Fig. 9a). We tested whether the induction of *CCA1* depends on *PRR7*. *CCA1:LUC* induction was significantly attenuated in *prp7-11* mutants compared with wild-type plants (Extended Data Fig. 9b). These data are consistent with sucrose activating *CCA1* through the repression of *PRR7*. Next, we examined whether *PRR7* contributes to the circadian period adjustment mediated by sucrose. Exogenous sucrose shortened the period of the circadian rhythm of *CCA1:LUC* activity by 2.7 h in wild-type plants, whereas the circadian rhythm of *CCA1:LUC* activity was not shortened in *prp7-11* mutants exposed to exogenous sucrose (Fig. 4a). Similarly, the period of the circadian rhythm of delayed fluorescence¹⁷ was also shortened by exogenous sucrose exposure in wild-type plants but not *prp7-11* mutants (Extended Data Fig. 9c).

To assess whether *PRR7* also has a role in circadian entrainment by sugars, we determined a PRC for *prp7-11* to pulses of exogenous sucrose. In contrast to wild-type plants (Fig. 2c), sucrose did not induce phase advances in *prp7-11* mutants (Extended Data Fig. 9d). Because SENSITIVE TO FREEZING 6 (*SFR6*), a subunit of the Mediator complex, contributes to the period adjustment mediated by sucrose (by an unknown mechanism)⁹, we determined whether other previously identified pathways participate in the regulation of the circadian oscillator by sugars. We measured rhythms in a range of circadian, sugar-insensitive and light-signalling mutants. With the exception of *cca1-11*, all of the tested mutants had a significantly shorter circadian period in

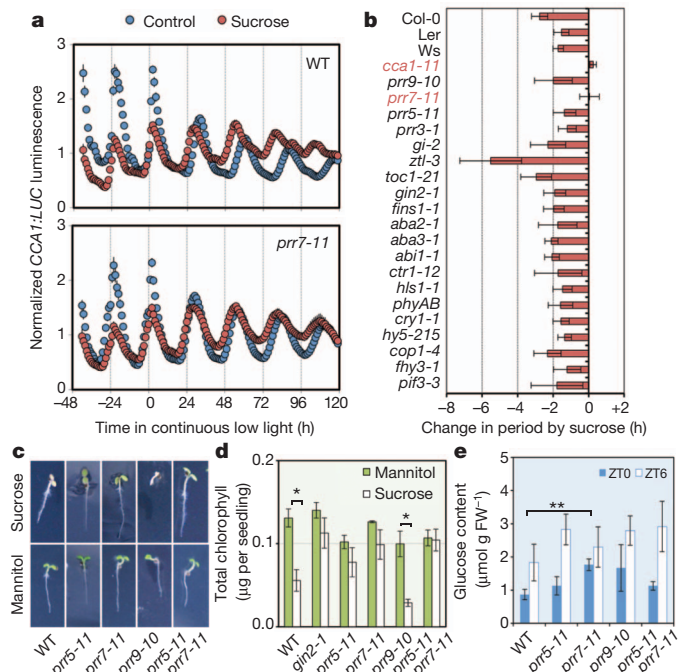


Figure 4 | *PRR7* contributes to circadian sugar signalling. **a**, *CCA1:LUC* activity rhythms in wild-type (WT) and *prp7-11* mutant seedlings in continuous low light in the presence or absence of exogenous sucrose (mean \pm s.e.m.; $n = 4$). **b**, Change in the period of *CCA1:LUC* activity rhythms in circadian, sugar-insensitive and light-signalling mutants, *CAB2:LUC* activity rhythms (W and *cca1-11*) and *CCR2:LUC* activity rhythms (*toc1-21*) grown in the presence of sucrose in continuous low light compared with seedlings grown in control media (mean \pm s.d.; $n = 8$). Red text indicates genotypes that were insensitive to sucrose. **c**, Seedlings germinated on 180 mM sucrose or 180 mM mannitol. Original magnification, 0.5 \times . **d**, Total chlorophyll content of seedlings germinated on 90 mM sucrose or 90 mM mannitol (mean \pm s.d.; $n = 3$). **e**, Glucose content of seedlings at dawn (mean \pm s.d.; $n = 3$). * $P < 0.05$; ** $P < 0.01$; two-tailed Student's *t*-test. *n* refers to number of biological replicates. FW, fresh weight.

the presence of sucrose than in control media (Fig. 4b and Extended Data Fig. 10). Together, these data indicate that *PRR7*, acting through *CCA1*, has a specific role in the regulation of the circadian clock by photosynthetically derived sugars and that this might occur through a novel signalling pathway.

Our findings led us to test whether *PRR7* might be more widely involved in sugar signalling. When germinated on medium containing 180 mM (6% w/v) sucrose, *prp7-11* mutants were resistant to the repression of chlorophyll accumulation (Fig. 4c, d), similar to *glucose insensitive 2 (hexokinase 1) (gin2-1)* mutants¹⁸. In addition, *prp7-11* seedlings had an elevated endogenous glucose concentration around dawn compared with wild-type seedlings (Fig. 4e), suggesting a role for *PRR7* in regulating endogenous sugar accumulation. This finding is consistent with previous reports of the involvement of *PRR* proteins in regulating chlorophyll biosynthesis and primary metabolism^{19–21}.

Altered feeding cycles can influence the phase of peripheral clocks in animals^{22,23}. Similarly, it has previously been suggested that a shoot-derived photosynthesis product might regulate a simplified circadian oscillator in *A. thaliana* roots²⁴. Photosynthesis also contributes to entrainment by an unknown mechanism in the green alga *Chlamydomonas reinhardtii*²⁵. By analysing the effects of altered photosynthesis products on free-running circadian rhythms and by examining the role of *PRR7*, we have demonstrated that photosynthetically derived sugars provide metabolic feedback that entrains the *A. thaliana* circadian clock in shoots. We propose that following light activation of *PRR7* at dawn, the accumulation of endogenous sugars as a result of photosynthesis represses the *PRR7* promoter, leading to de-repression of *CCA1*.

Thus, *PRR7* expression is coordinately modulated by light and photosynthesis, allowing *PRR7* to function as a transcriptional repressor in circadian sugar signalling (Extended Data Fig. 1). Our results define a novel metabolic feedback loop that contributes to circadian entrainment in plants.

METHODS SUMMARY

Growth conditions and media. Surface-sterilized *A. thaliana* seeds (on Col-0, Ler and Ws backgrounds) were sown on half-strength Murashige and Skoog agar medium (1/2 MS) without sucrose. After 2 days at 4 °C in darkness, the seeds were grown at 19 °C in a 12 h light–12 h dark cycle in 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ white fluorescent light. Chemicals were added to the media as follows: 90 mM sugars, 20 μM DCMU, 5 μM norflurazon and 220 $\mu\text{g ml}^{-1}$ lincomycin. For CO_2 -free air, ambient air was pumped through soda lime. Sugars were added topically to dark-adapted seedlings to a final concentration of approximately 30 mM in the media, unless otherwise indicated. For gating and PRC experiments, the seedlings were germinated on 1 μm pore-diameter nylon mesh, and pulses were applied by transferring the seedlings to media containing 90 mM sucrose or 90 mM mannitol for 3 h. PRCs were calculated as described previously¹⁴.

Luciferase experiments. LUC reporter lines were grown in clusters of 5–10 seedlings, supplied with 1–2 mM D-luciferin and released into continuous light or dark conditions after 7–11 days in cycles of 12 h light and 12 h dark. Seedlings were transferred to treatments 48–60 h before release into continuous light or dark conditions. Light was supplied from red- and blue-light-emitting diodes (LEDs) at 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (high light) or 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (low light). Circadian period and relative amplitude error estimates were determined using the Biological Rhythms Analysis Software System (BRASS). All *n* values represent biological replicates, and all data are representative of independently repeated experiments.

Real-time PCR. RNA was extracted from three biological replicates of 12-day-old seedlings growing in a cycle of 12 h light and 12 h dark from 37–58 h after transfer to treatments at dusk. cDNA was synthesized from DNase-treated RNA with oligo(dT) primers. Gene-specific primers (see Methods) were used to amplify products on a Rotor-Gene 6000 Real-Time PCR machine.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 February; accepted 23 August 2013.

Published online 23 October 2013.

- Woelfle, M. A., Ouyang, Y., Phanvijhitsiri, K. & Johnson, C. H. The adaptive value of circadian clocks?: an experimental assessment in Cyanobacteria. *Curr. Biol.* **14**, 1481–1486 (2004).
- Dodd, A. N. *et al.* Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* **309**, 630–633 (2005).
- Harmer, S. L. The circadian system in higher plants. *Annu. Rev. Plant Biol.* **60**, 357–377 (2009).
- Bass, J. Circadian topology of metabolism. *Nature* **491**, 348–356 (2012).
- Harmer, S. L. *et al.* Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**, 2110–2113 (2000).
- Lu, Y., Gehan, J. P. & Sharkey, T. D. Daylength and circadian effects on starch degradation and maltose metabolism. *Plant Physiol.* **138**, 2280–2291 (2005).
- Graf, A., Schlereth, A., Stitt, M. & Smith, A. M. Circadian control of carbohydrate availability for growth in *Arabidopsis* plants at night. *Proc. Natl Acad. Sci. USA* **107**, 9458–9463 (2010).
- Noordally, Z. B. *et al.* Circadian control of chloroplast transcription by a nuclear-encoded timing signal. *Science* **339**, 1316–1319 (2013).
- Knight, H., Thomson, A. J. W. & McWatters, H. G. SENSITIVE TO FREEZING6 integrates cellular and environmental inputs to the plant circadian clock. *Plant Physiol.* **148**, 293–303 (2008).
- Dalchau, N. *et al.* The circadian oscillator gene *GIGANTEA* mediates a long-term response of the *Arabidopsis thaliana* circadian clock to sucrose. *Proc. Natl Acad. Sci. USA* **108**, 5104–5109 (2011).
- Bläsing, O. E. *et al.* Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell* **17**, 3257–3281 (2005).
- Koussevitzky, S. *et al.* Signals from chloroplasts converge to regulate nuclear gene expression. *Science* **316**, 715–719 (2007).
- Lai, A. G. *et al.* CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proc. Natl Acad. Sci. USA* **109**, 17129–17134 (2012).
- Johnson, C. H. in *Circadian Clocks from Cell to Human* (eds Hiroshige, T. & Honma, K.) 209–249 (Hokkaido Univ. Press, 1992).
- Michael, T. P., Salomé, P. A. & McClung, C. R. Two *Arabidopsis* circadian oscillators can be distinguished by differential temperature sensitivity. *Proc. Natl Acad. Sci. USA* **100**, 6878–6883 (2003).
- Nakamichi, N. *et al.* PSEUDO-RESPONSE REGULATORS 9, 7 and 5 are transcriptional repressors in the *Arabidopsis* circadian clock. *Plant Cell* **22**, 594–605 (2010).
- Gould, P. D. *et al.* Delayed fluorescence as a universal tool for the measurement of circadian rhythms in higher plants. *Plant J.* **58**, 893–901 (2009).
- Moore, B. *et al.* Role of the *Arabidopsis* glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* **300**, 332–336 (2003).
- Kato, T. *et al.* Mutants of circadian-associated PRR genes display a novel and visible phenotype as to light responses during de-etiolation of *Arabidopsis thaliana* seedlings. *Biosci. Biotechnol. Biochem.* **71**, 834–839 (2007).
- Fukushima, A. *et al.* Impact of clock-associated *Arabidopsis* pseudo-response regulators in metabolic coordination. *Proc. Natl Acad. Sci. USA* **106**, 7251–7256 (2009).
- Nakamichi, N. *et al.* Transcriptional repressor PRR5 directly regulates clock-output pathways. *Proc. Natl Acad. Sci. USA* **109**, 17123–17128 (2012).
- Damiola, F. *et al.* Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev.* **14**, 2950–2961 (2000).
- Stokkan, K.-A., Yamazaki, S., Tei, H., Sakaki, Y. & Menaker, M. Entrainment of the circadian clock in the liver by feeding. *Science* **291**, 490–493 (2001).
- James, A. B. *et al.* The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science* **322**, 1832–1835 (2008).
- Johnson, C. H., Kondo, T. & Hastings, J. W. Action spectrum for resetting the circadian phototaxis rhythm in the CW15 strain of *Chlamydomonas*. *Plant Physiol.* **97**, 1122–1129 (1991).

Acknowledgements This work was supported by BBSRC grant BB/H006826/1. We thank J. O'Neill, J. Davies and J. Hibberd for comments on the manuscript.

Author Contributions M.J.H. and A.A.R.W. designed the research. M.J.H., O.M., F.C.R. and K.E.H. performed the experiments and analysed the data. M.J.H. and A.A.R.W. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.A.R.W. (aarw2@cam.ac.uk).

METHODS

Plant materials and growth methods. *A. thaliana* of the Col-0 ecotype was transformed with *CCA1:LUC*, *TOC1:LUC*, *PRR7:LUC* or *CCR2:LUC*. *A. thaliana* of the Ws ecotype was transformed with *GI:LUC*, *PRR9:LUC* or *CAB2:LUC*. *CCA1:LUC* was introduced into the Ler ecotype and the following mutants by crossing: *prp5-11*, *prp7-11*, *prp9-10* (ref. 26), *prp3-1* (ref. 27), *gigantea* (*gi-2*)²⁸, *zeittlupe* (*ztl-3*)²⁹, *gin2-1* (ref. 18), *fructose-insensitive-1* (*fins1-1*)³⁰, *abscisic acid deficient 2* (*aba2-1*; *gin1*)³¹, *aba3-1* (*gin5*) (ref. 32), *abscisic acid insensitive 1* (*abi1-1*)³³, *constitutive triple response 1* (*ctr1-12*; *gin4*)³⁴, *hookless-1-1* (*hls1-1*)³⁵, *phytochrome A phytochrome B* (*phyA-201*; *phyB-5*)³⁶, *cryptochrome 1* (*cry1-1*)³⁷, *long hypocotyl 5* (*hy5-215*)³⁸, *constitutive photomorphogenic 1* (*cop1-4*)³⁹, *far-red elongated hypocotyl 3* (*fhy3-1*)⁴⁰, *phytochrome interacting factor 3* (*pif3-3*)⁴¹ and *spindly* (*spy-3*)⁴². Ler/*CCA1:LUC* and *gin2-1/CCA1:LUC* were backcrossed to Ler or *CCA1:LUC*, respectively, twice. The *gin2-1*(Col-0) line was used for sugar-insensitivity experiments to allow direct comparison with *prp* mutants and Col-0. Surface-sterilized seeds were sown on half-strength Murashige and Skoog media (1/2 MS), pH 5.7, without sucrose and solidified with 0.8% (w/v) Bacto agar. After sowing, seeds were kept at 4 °C in darkness for 2 days, then grown in 12 h light–12 h dark cycles under 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ cool fluorescent white light at a constant temperature of 19 °C.

Photon-counting experiments. Clusters of five to ten seedlings were grown on 1/2 MS agar medium and entrained in 12 h light and 12 h dark cycles (50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light). For LUC measurement, seedlings were dosed twice with 1–2 mM D-luciferin 12–48 h before commencing photon counting. Seedlings were released into continuous light after 7–11 days of light–dark cycles. Luminescence was detected for 800 s at each time point with an HRPC54 (Photek) or an LB 985 NightSHADE (Berthold) camera. Delayed fluorescence⁴⁷ was measured for 5 s with an LB 985 NightSHADE camera. During photon counting, light was supplied from red (660 nm) and blue (470 nm) light-emitting diodes (LEDs) at 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ during light–dark cycles and either 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (continuous light) or 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (continuous low light) during continuous light conditions. Where indicated, data for each replicate were normalized to average counts across the experiment. All period and relative amplitude error estimates were performed on rhythms for 24–120 h in continuous conditions on non-normalized data using fast Fourier Transform–nonlinear least squares (FFT–NLLS) analysis, implemented in the Biological Rhythms Analysis Software System (BRASS) (<http://millar.bio.ed.ac.uk/PEBrown/BRASS/BrassPage.htm>). All *n* values represent biological replicates, and all data are representative of independently repeated experiments. Tests were justified by determining the minimum difference of means with a power of 0.9. Two-sided statistical tests, including assessments of normal distribution and equal variance, were performed in Microsoft Excel.

The gating of short-term responses of LUC reporters to sugars at 1.5 h intervals for 24 h was performed in 8-day-old seedlings from ZT0 in continuous low light. The signal was normalized to the average signal across a time course of several days. The change in normalized LUC reporter activity in seedlings before and after 3 h exposure to 90 mM sugars was subtracted from the change in normalized LUC reporter activity in untreated seedlings.

PRC experiments were performed in 8-day-old seedlings from ZT0 in continuous low light. Seedlings growing on 1 μm pore-diameter nylon mesh on 1/2 MS were transferred to 1/2 MS containing 90 mM sugars for 3 h at 1.5 h intervals. The phase was determined based on the time of the circadian peaks following sugar pulses, and the PRC was determined relative to the phase in control seedlings after accounting for period differences, as described previously¹⁴.

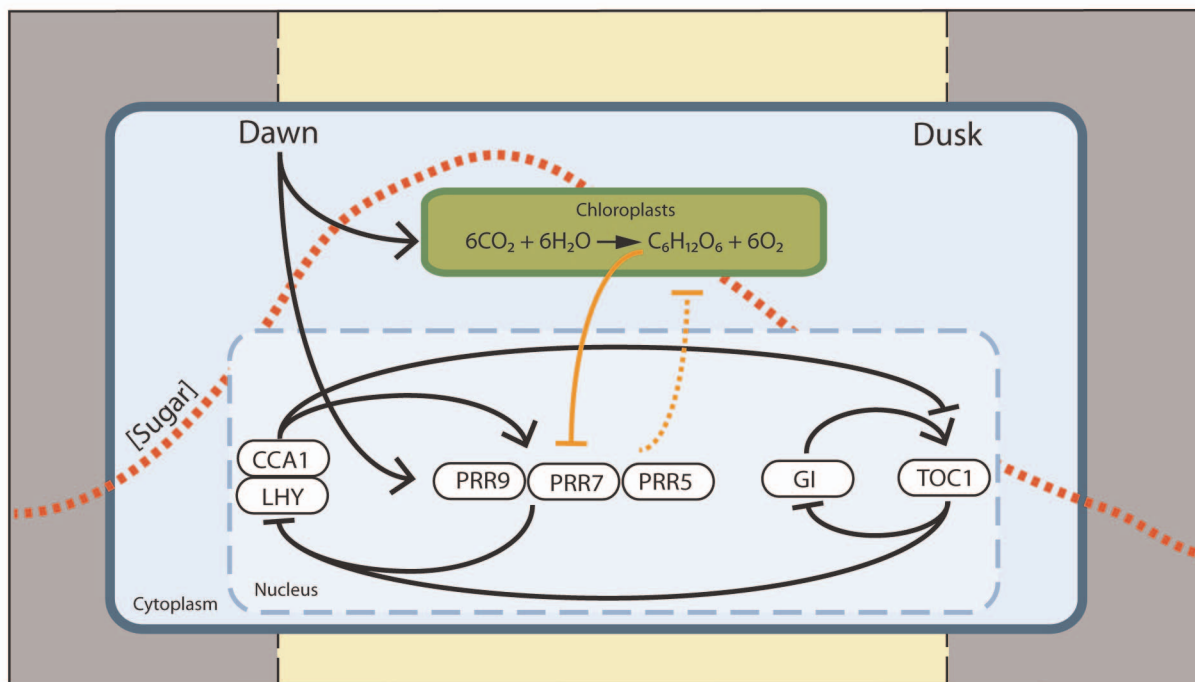
Treatments. Sugars were added to the media to a final concentration of 90 mM (3% w/v) sucrose unless otherwise indicated. Chemical treatments were added to the media at the following concentrations: 20 μM DCMU, 5 μM norflurazon and 220 $\mu\text{g ml}^{-1}$ lincomycin. Seedlings were transferred to treatments 48–60 h before release into continuous conditions. CO₂-free air was produced by pumping ambient air through self-indicating soda lime, a 0.45 μm filter and autoclaved deionized water into a sealed growth plate with an outlet. The CO₂ concentration of the air from the outlet was confirmed at <1 p.p.m. using an infrared gas analyser (ADC 255-MK3). For experiments with dark-adapted seedlings, sugars were added with a micropipette to about 0.1 volume of the growth medium, to give a final concentration of ~30 mM. For gating and PRC experiments, the seedlings were transferred to media containing 90 mM sugars. Treatments in the dark were performed under dim green light.

Real-time PCR. Ten-day-old seedlings growing in 12 h light–12 h dark cycles (50 $\mu\text{mol m}^{-2} \text{s}^{-1}$) were transferred to treatments at dusk, and leaf tissue was

collected at 3 h intervals 37–58 h later. Total RNA was extracted from three biological replicates of frozen leaf tissue using the RNeasy Plant Mini Kit (QIAGEN) and RNase-Free DNase on-column treatment (QIAGEN). cDNA was synthesized from 1 μg RNA with the RevertAid First Strand cDNA Synthesis Kit (Fermentas) using oligo(dT) primers. Technical replicates of gene-specific products were amplified in 10 μl reactions using the Rotor-Gene SYBR Green PCR Kit on a Rotor-Gene 6000 Real-Time PCR machine fitted with a Rotor-Disc 100 (QIAGEN). The primer sequences were as follows: *UBQ10-F*, 5'-GGCCTTGATAATCCCTGATGAATAAG-3'; *UBQ10-R*, 5'-AAAGAGATAACAGGAACGGAACATAGT-3'; *CCA1-F*, 5'-GATGATGTTGAGCGGATG-3'; *CCA1-R*, 5'-TGGTGTTAACTGAGCTGTGAAG-3'; *LHY-F*, 5'-ACGAACAGGTAAGTGGCGACATT-3'; *LHY-R*, 5'-TGGGAACATCTTGAACCGCTT-3'; *PRR9-F*, 5'-CCACAGTAACGAATCAGAAGCAA-3'; *PRR9-R*, 5'-TTGTCCAGCAATCCCTCA-3'; *PRR7-F*, 5'-GGAACTTGGCGGATGAAAA-3'; *PRR7-R*, 5'-CGAGGGCGTGTCTCTGCT-3'; *PRR5-F*, 5'-CCGAATGAAGCGAAAGGACA-3'; *PRR5-R*, 5'-GGATTGACTTGACGAACG-3'. Relative transcript levels were determined by incorporating PCR efficiencies as described previously⁴³.

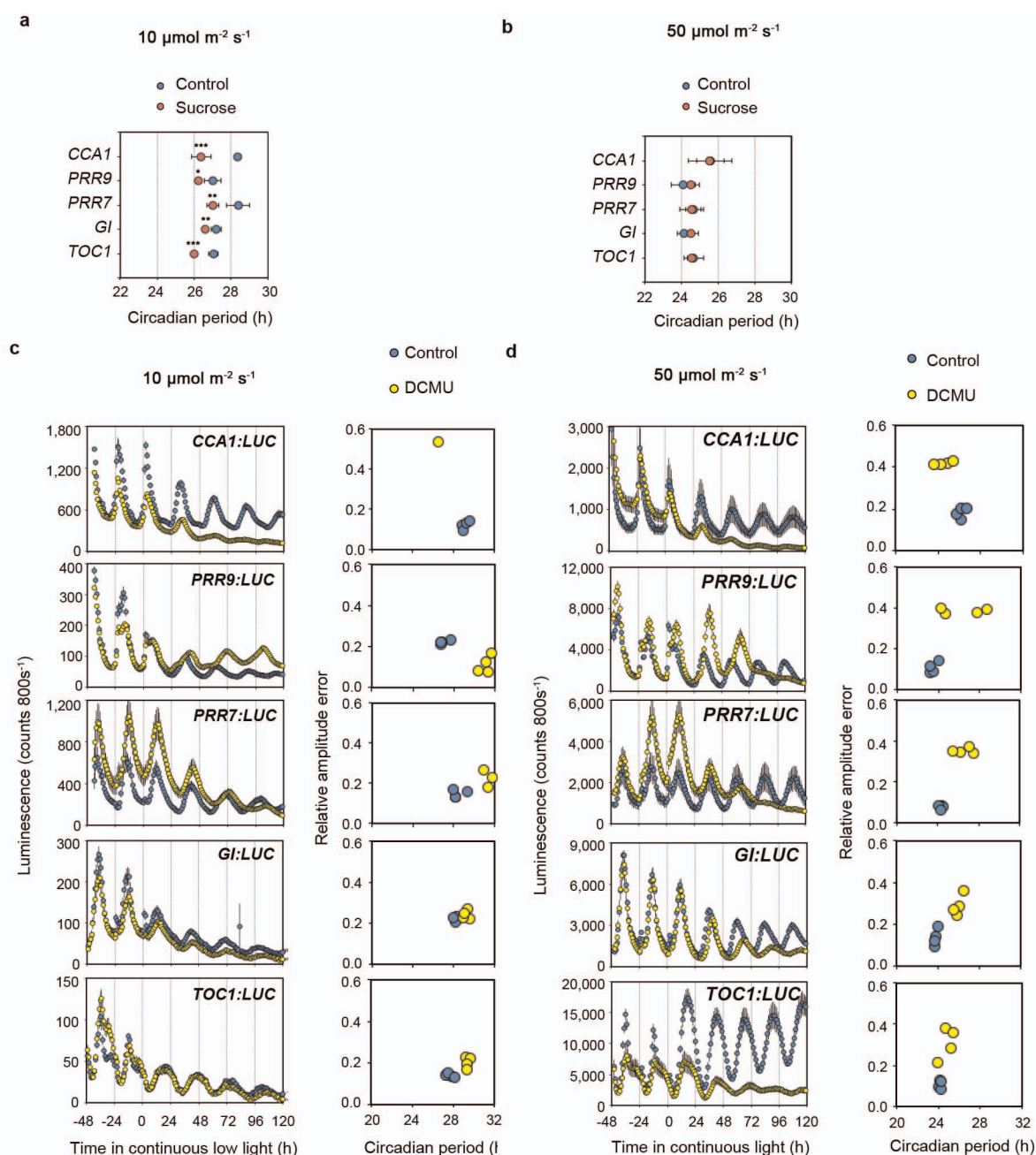
Sugar and chlorophyll measurements. For soluble sugar measurements, 50–100 mg frozen tissue was extracted twice in 80% (v/v) ethanol and used immediately to determine sugar concentrations with the Sucrose/Fructose/D-Glucose assay kit (Megazyme). For chlorophyll measurements, fresh tissue was extracted in methanol, and the concentrations were determined as described previously⁴⁴.

26. Nakamichi, N., Kita, M., Ito, S., Yamashino, T. & Mizuno, T. PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, together play essential roles close to the circadian clock of *Arabidopsis thaliana*. *Plant Cell Physiol.* **46**, 686–698 (2005).
27. Para, A. et al. PRR3 is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell* **19**, 3462–3473 (2007).
28. Park, D. H. Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis* GIGANTEA gene. *Science* **285**, 1579–1582 (1999).
29. Jarillo, J. A. et al. An *Arabidopsis* circadian clock component interacts with both CRY1 and phyB. *Nature* **410**, 487–490 (2001).
30. Cho, Y.-H. & Yoo, S.-D. Signaling role of fructose mediated by FINS1/FBP in *Arabidopsis thaliana*. *PLoS Genet.* **7**, e1001263 (2011).
31. González-Guzmán, M. et al. The short-chain alcohol dehydrogenase ABA2 catalyzes the conversion of xanthoxin to abscisic aldehyde. *Plant Cell* **14**, 1833–1846 (2002).
32. Xiong, L., Ishitani, M., Lee, H. & Zhu, J. K. The *Arabidopsis* LOS5/ABA3 locus encodes a molybdenum cofactor sulfurylase and modulates cold stress- and osmotic stress-responsive gene expression. *Plant Cell* **13**, 2063–2083 (2001).
33. Leung, J. et al. *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phosphatase. *Science* **264**, 1448–1452 (1994).
34. Gibson, S. I., Laby, R. J. & Kim, D. The sugar-insensitive1 (*sis1*) mutant of *Arabidopsis* is allelic to *ctr1*. *Biochem. Biophys. Res. Commun.* **280**, 196–203 (2001).
35. Lehman, A., Black, R. & Ecker, J. R. *HOOKLESS1*, an ethylene response gene, is required for differential cell elongation in the *Arabidopsis* hypocotyl. *Cell* **85**, 183–194 (1996).
36. Reed, J. W., Nagatani, A., Elich, T. D., Fagan, M. & Chory, J. Phytochrome A and phytochrome B have overlapping but distinct functions in *Arabidopsis* development. *Plant Physiol.* **104**, 1139–1149 (1994).
37. Ahmad, M. & Cashmore, A. R. *HY4* gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor. *Nature* **366**, 162–166 (1993).
38. Oyama, T., Shismura, Y. & Okada, K. The *Arabidopsis* *HY5* gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes Dev.* **11**, 2983–2995 (1997).
39. Deng, X. W. et al. *COP1*, an *Arabidopsis* regulatory gene, encodes a protein with both a zinc-binding motif and a G_p homologous domain. *Cell* **71**, 791–801 (1992).
40. Hudson, M. E., Lisch, D. R. & Quail, P. H. The *FHY3* and *FAR1* genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. *Plant J.* **34**, 453–471 (2003).
41. Ni, M., Tepperman, J. M. & Quail, P. H. PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell* **95**, 657–667 (1998).
42. Jacobsen, S. E., Binkowski, K. A. & Olszewski, N. E. SPINDLY, a tetratricopeptide repeat protein involved in gibberellin signal transduction in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **93**, 9292–9296 (1996).
43. Talke, I. N., Hanikenne, M. & Krämer, U. Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator *Arabidopsis halleri*. *Plant Physiol.* **142**, 148–167 (2006).
44. Porra, R. J., Thompson, W. A. & Kreidemann, P. E. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta* **975**, 384–394 (1989).
45. Kwak, J. M. et al. NADPH oxidase *AtrbohD* and *AtrbohF* genes function in ROS-dependent ABA signaling in *Arabidopsis*. *EMBO J.* **22**, 2623–2633 (2003).



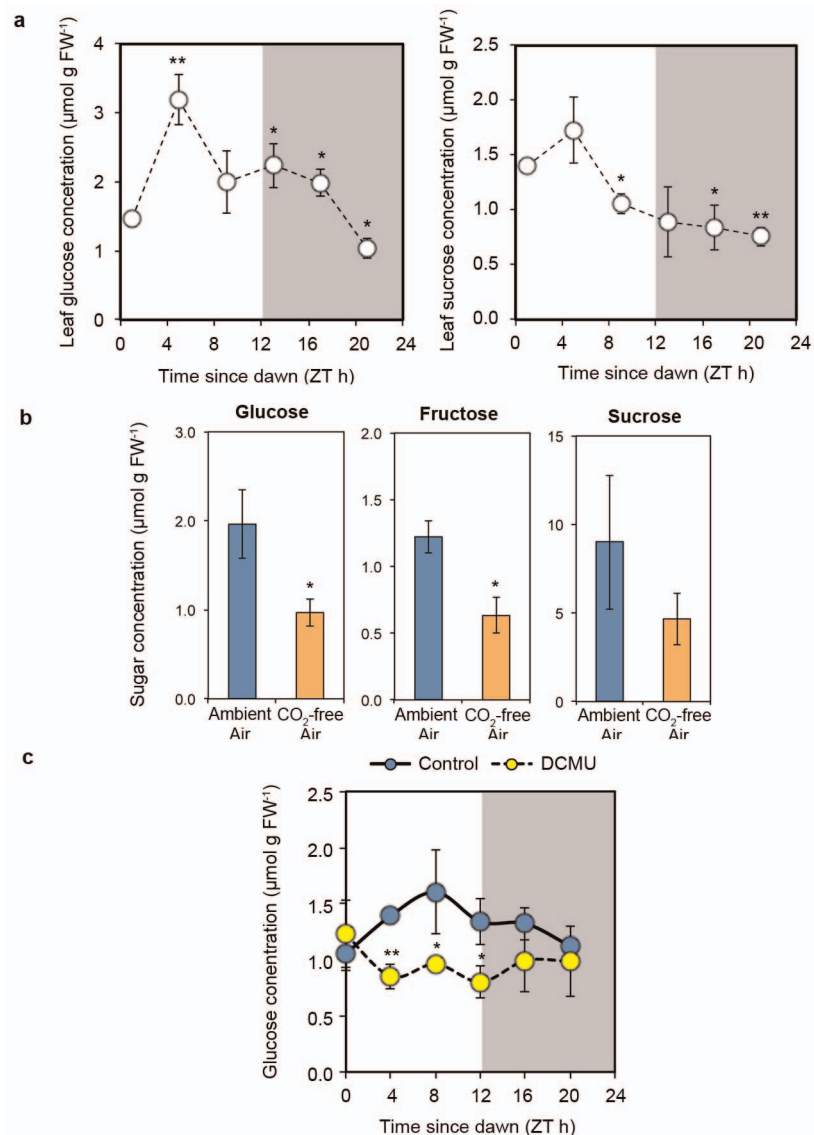
Extended Data Figure 1 | A model for entrainment of the *A. thaliana* circadian clock by photosynthetically derived sugars. From dawn, light activates *PRR7* and drives photosynthesis. The concentrations of simple sugars produced by photosynthesis accumulate within the plant during the day (red dashed line), peaking around 4–8 h after dawn. High endogenous sugar concentrations lead to suppression of the *PRR7* promoter, contributing to the phase of *PRR7* rhythms. *PRR7* is a transcriptional repressor of the circadian clock component *CCA1*. Thus, the rhythms of endogenous sugars derived from

photosynthesis contribute to circadian entrainment through *PRR7*. We propose that the timing of these events represents a 'metabolic dawn'. Dawn is a time-dependent gradient of light intensity, whereas 'metabolic dawn' represents a gradient of increasing metabolite concentration. The metabolic gradient lags behind that of light and contributes to the setting of the circadian clock. In the model, previously established relationships are shown by black connectors, and novel relationships proposed in this study are shown by orange connectors.



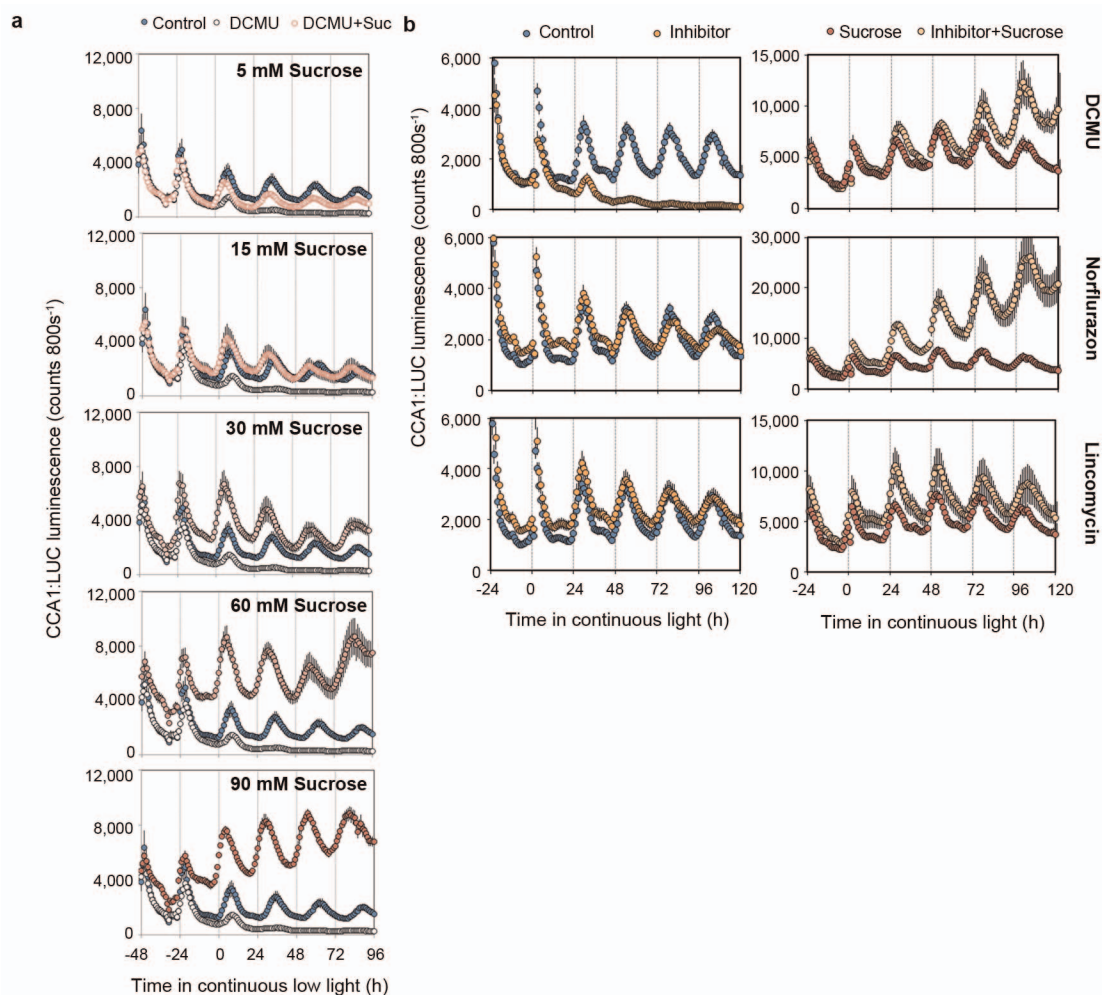
Extended Data Figure 2 | Effects of exogenous sucrose and photosynthesis inhibition on circadian rhythms. **a, b,** Period estimates for the rhythms of the promoter:LUC reporters in continuous low light (**a**) or continuous light (**b**) in plants grown in media with or without sucrose (mean \pm s.d.; $n = 4$). **c, d,** Promoter:LUC reporter rhythms (mean \pm s.e.m.) and relative amplitude

error versus period plots for seedlings in media in the presence or absence of DCMU in continuous low light (**c**) or continuous light (**d**) ($n = 4$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; compared with untreated plants by two-tailed Student's t -test. n refers to number of biological replicates.



Extended Data Figure 3 | The rhythms of endogenous sugars peak in the morning and are reduced by inhibition of photosynthesis. **a**, Leaf sucrose and glucose concentrations in 10-day-old seedlings growing in 12 h light and 12 h dark cycles (mean \pm s.d.; $n = 3$). **b**, Glucose, fructose and sucrose concentrations 4 h after subjective dawn in 13-day-old seedlings grown in CO_2 -free air or ambient air in continuous low light for 5 days (mean \pm s.d.; $n = 3$).

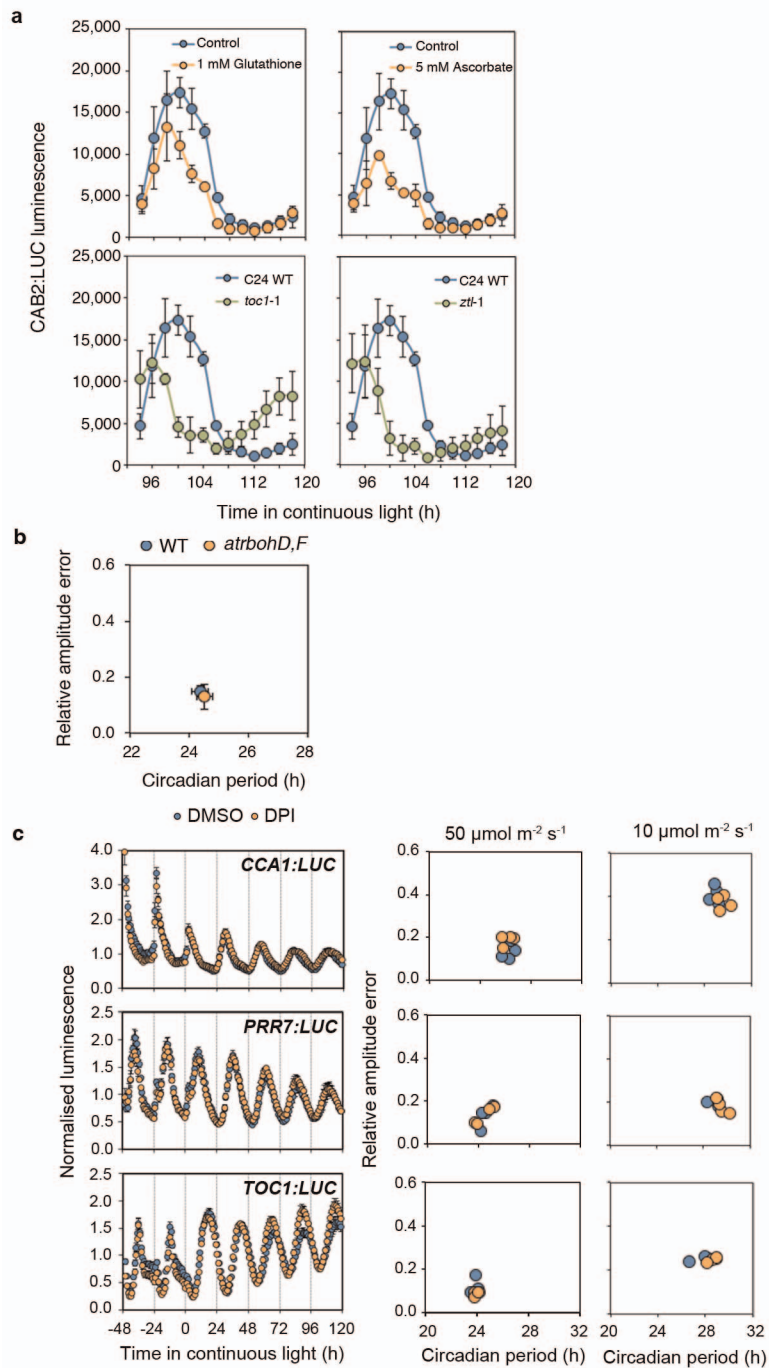
c, Glucose concentration in 10-day-old seedlings growing in a light–dark cycle at 12–36 h after transfer to DCMU or control media at dusk (mean \pm s.d.; $n = 3$). * $P < 0.05$; ** $P < 0.01$ by two-tailed Student's t -test compared with ZT0 in **a** and compared with control conditions in **b** and **c**. n refers to number of biological replicates.



Extended Data Figure 4 | Effects of DCMU, norflurazon or lincomycin on *CCA1:LUC* rhythms in the presence or absence of exogenous sucrose.

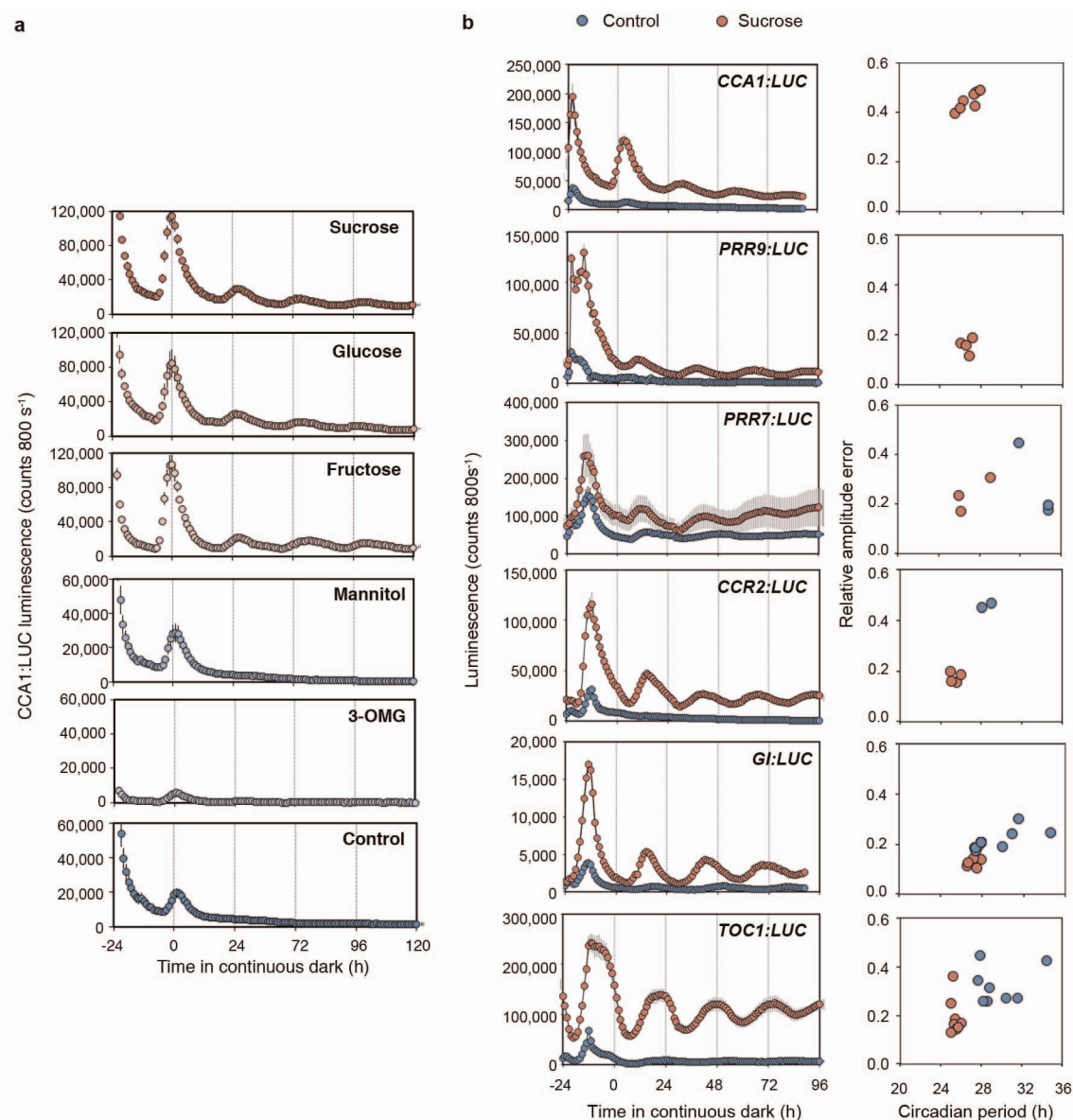
a, *CCA1:LUC* rhythms in continuous low light for seedlings transferred to media containing DCMU in the presence of the indicated exogenous sucrose concentrations compared with control media (mean \pm s.e.m.; $n = 4$).

b, *CCA1:LUC* rhythms in continuous light for seedlings transferred to media containing DCMU, norflurazon or lincomycin in the absence (left) or presence (right) of exogenous sucrose (mean \pm s.e.m.; $n = 4$). n refers to number of biological replicates.



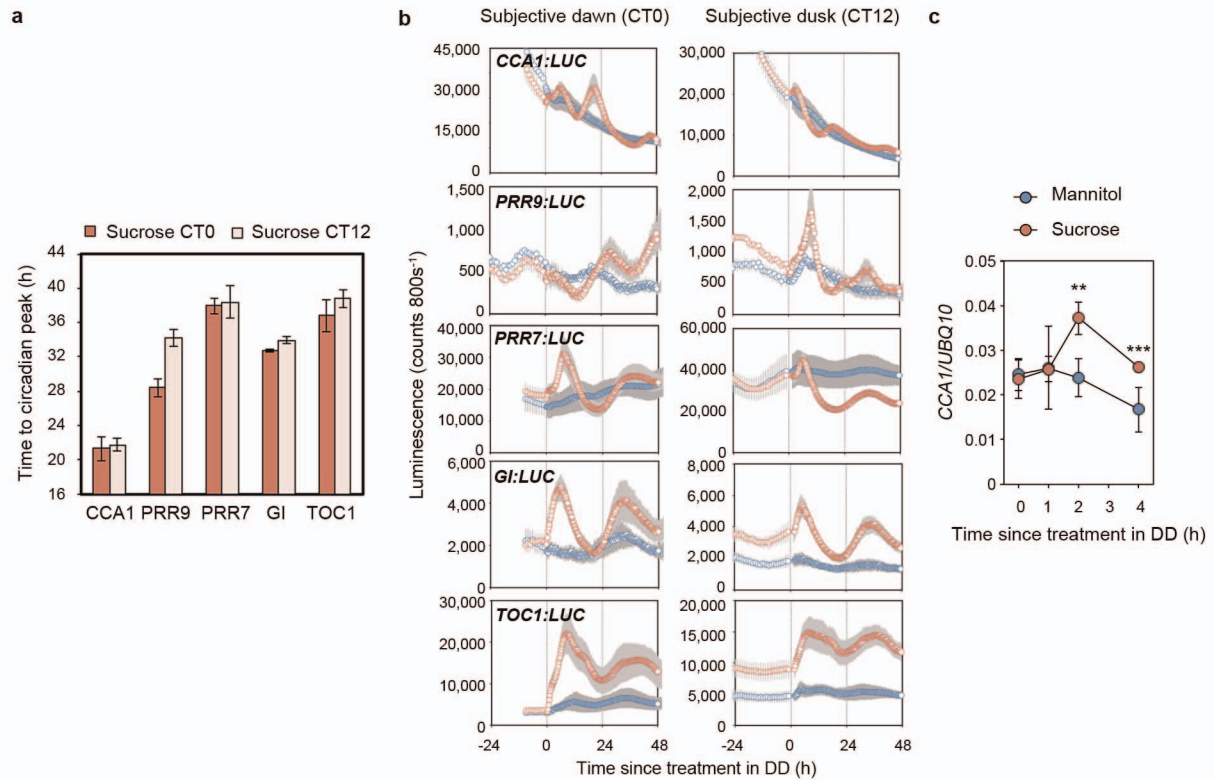
Extended Data Figure 5 | Altering reactive oxygen species (ROS) production does not influence circadian rhythms. **a**, CAB2:LUC rhythms in seedlings transferred to continuous light and treated with 1 mM glutathione or 5 mM ascorbate. The short-period mutant *toc1-1* and long-period mutant *ztl-1* were included as positive controls (means \pm s.d.; $n = 2-3$). **b**, Relative amplitude error versus period plot for leaf movement rhythms in wild-type

plants and NADPH oxidase *rbbohD,F* mutants⁴⁵ in continuous light (mean \pm s.e.m.). **c**, Promoter:LUC rhythms and relative amplitude error versus period plots for seedlings grown in continuous light or continuous low light and treated with 10 μM diphenyleneiodonium (DPI) or 0.1% (v/v) dimethylsulphoxide (DMSO) at 0 h (mean \pm s.e.m.; $n = 4$). n refers to number of biological replicates.



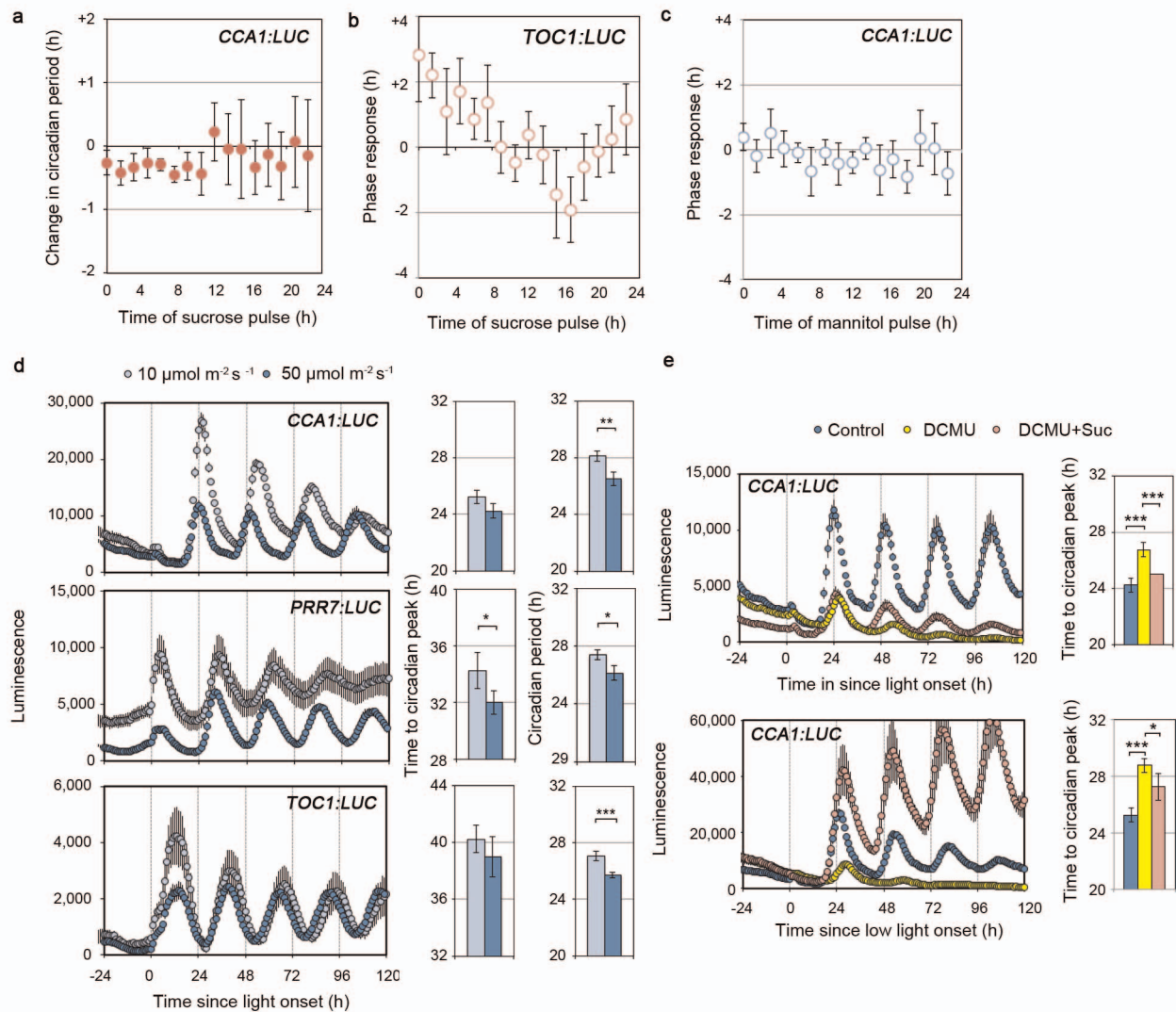
Extended Data Figure 6 | Metabolically active sugars sustain circadian rhythms in darkness. **a**, *CCA1:LUC* rhythms in continuous dark in seedlings grown in media containing the indicated sugars or control treatments (mean \pm s.e.m.; $n = 4$). **b**, Promoter:LUC rhythms (mean \pm s.e.m.; $n = 4$) and relative amplitude error versus period plots ($n = 4$ –8) for seedlings grown in

continuous dark in media with or without sucrose. Note that rhythms could not be detected in seedlings grown without sucrose for the morning-expressed *CCA1:LUC* or *PRR9:LUC* but could be detected for the evening-expressed *GI:LUC* and *TOC1:LUC*, despite the small amplitude. n refers to number of biological replicates.



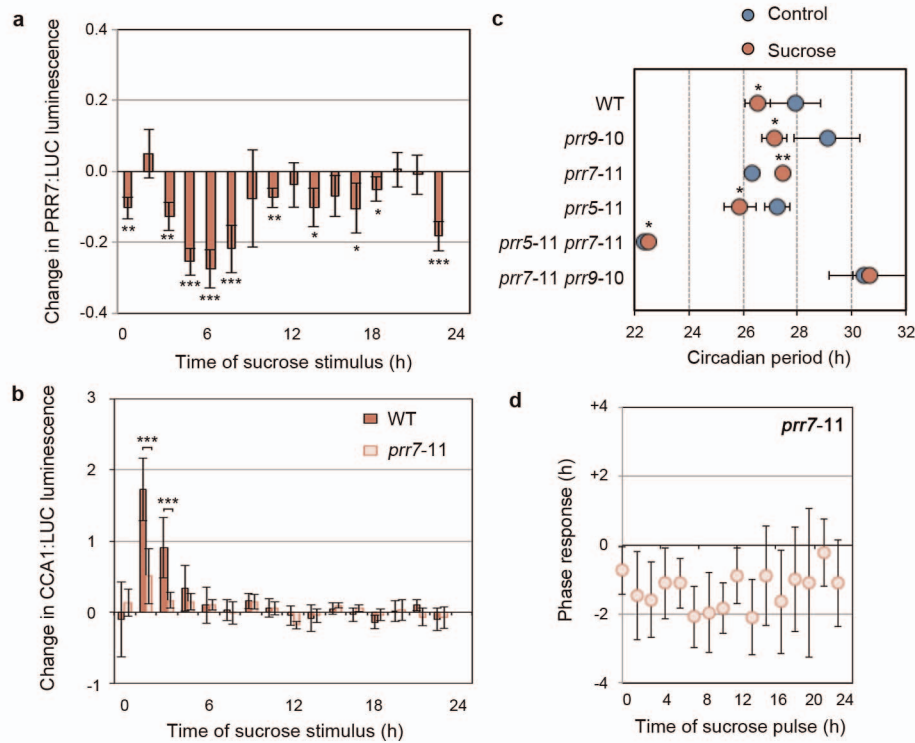
Extended Data Figure 7 | Exogenous sugar can set the circadian phase in dark-adapted seedlings. **a**, Time to the first circadian peak of promoter:LUC reporters in seedlings treated with sucrose after 72 h (subjective dawn, CT0) or 84 h (subjective dusk, CT12) in continuous dark (mean \pm s.d.; $n = 4$). **b**, Promoter:LUC rhythms of seedlings after sucrose or mannitol treatment as

in **a** (mean \pm s.e.m.; $n = 4$). **c**, *CCA1* transcript level relative to *UBQ10* in seedlings treated with sucrose or mannitol after 72 h in continuous dark (mean \pm s.d.; $n = 3$). ** $P < 0.01$; *** $P < 0.001$ by two-tailed Student's *t*-test. *n* refers to number of biological replicates.



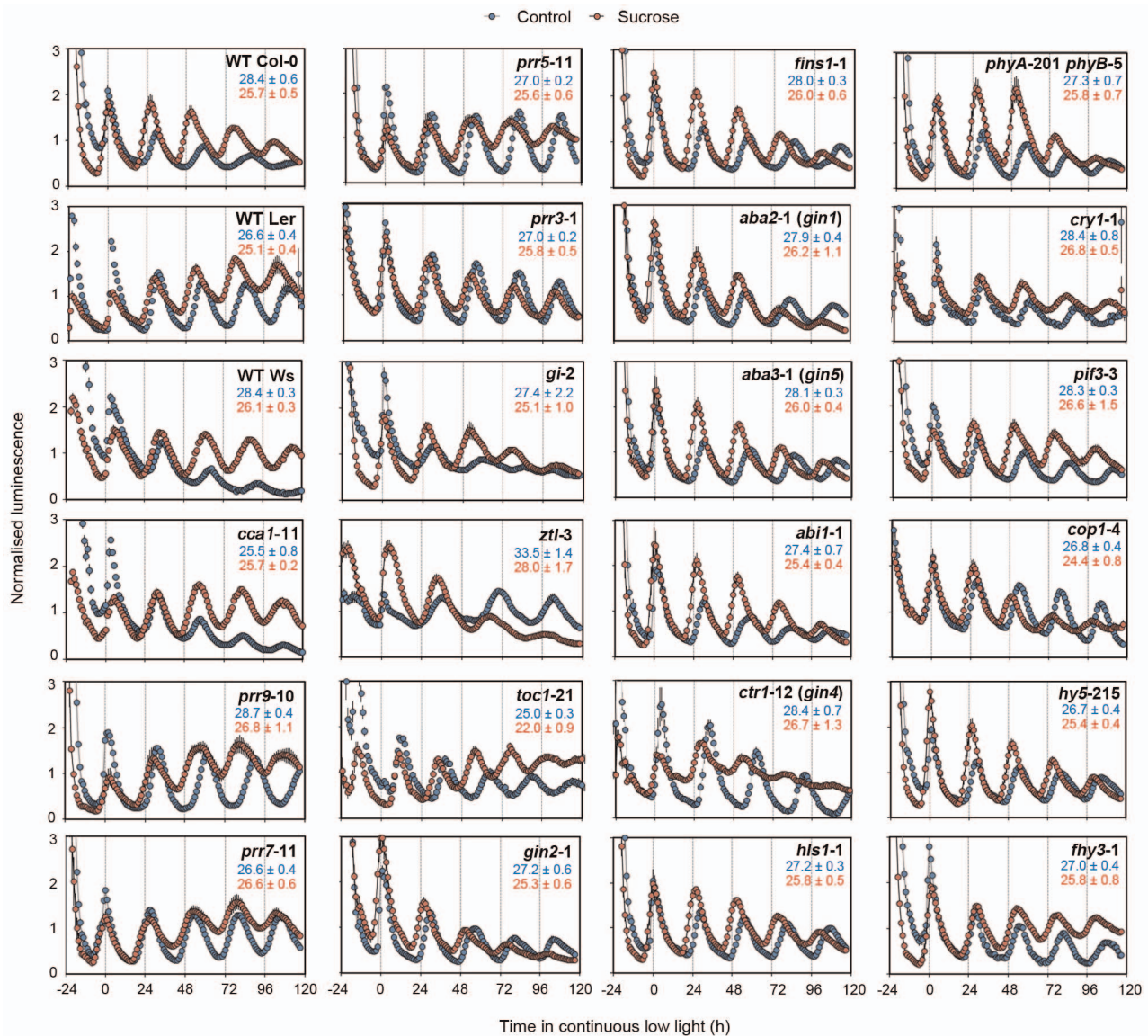
Extended Data Figure 8 | Phase setting by sugar and light. **a**, Change in the period of *CCA1:LUC* after pulses of sucrose compared with control seedlings in continuous low light (mean \pm s.d.; $n = 8$). **b**, Phase response of *TOC1:LUC* to pulses of sucrose for seedlings in continuous low light (mean \pm s.d.; $n = 8$). **c**, Phase response of *CCA1:LUC* to pulses of mannitol (mean \pm s.d.; $n = 8$). **d**, LUC reporter rhythms (mean \pm s.e.m.), time to the circadian peak (mean \pm s.d.) and period estimates (mean \pm s.d.) in seedlings grown in

continuous darkness for 72 h then transferred to continuous light or continuous low light ($n = 4$). **e**, *CCA1:LUC* rhythms (mean \pm s.e.m.) and time to the circadian peak in seedlings following transfer to continuous light or continuous low light in control media, medium containing DCMU, or medium containing DCMU and sucrose after 72 h in continuous dark ($n = 4$). * $P < 0.05$; *** $P < 0.001$ by two-tailed Student's t -test. n refers to number of biological replicates.



Extended Data Figure 9 | Regulation of the circadian clock by sugar requires PRR7. **a**, Change in *PRR7:LUC* luminescence after 3 h treatment with sucrose relative to untreated plants (mean \pm s.d.; $n = 4$). The data were normalized across the time series, and the change relative to untreated plants was plotted. **b**, Change in *CCA1:LUC* luminescence in wild-type plants and *prr7-11* mutants after 3 h treatment with sucrose relative to untreated plants (mean \pm s.d.; $n = 8$). The data were normalized across the time series, and the change relative to untreated plants of the appropriate genotype was plotted.

c, Period estimates of rhythms of delayed fluorescence in wild-type and mutant seedlings in continuous low light in media with or without exogenous sucrose (mean \pm s.d.; $n = 4$). **d**, Phase response of *CCA1:LUC* to pulses of sucrose in *prr7-11* seedlings in continuous low light (mean \pm s.d.; $n = 8$). Compare this with the sucrose PRC for *CCA1:LUC* in wild-type seedlings in Fig. 2c. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ by Student's two-tailed t -test compared with controls in **a** and **c** and compared with wild-type plants in **b**. n refers to number of biological replicates.



Extended Data Figure 10 | Effect of exogenous sucrose on circadian period in circadian, sugar-insensitive and light-signalling mutants. LUC reporter rhythms in circadian, sugar-insensitive and light-signalling mutants in continuous low light in media with or without exogenous sucrose

(mean ± s.e.m.; $n = 4$). The reporter is *CCA1:LUC* in all lines except for Ws, *cca1-11* (*CAB2:LUC*) and *toc1-21* (*CCR2:LUC*). Period estimates are shown in blue (control) and red (sucrose) for each line (mean ± s.d.; $n = 8$). n refers to number of biological replicates.

Synthetic non-oxidative glycolysis enables complete carbon conservation

Igor W. Bogorad^{1,2}, Tzu-Shyang Lin¹ & James C. Liao^{1,3}

Glycolysis, or its variations, is a fundamental metabolic pathway in life that functions in almost all organisms to decompose external or intracellular sugars. The pathway involves the partial oxidation and splitting of sugars to pyruvate, which in turn is decarboxylated to produce acetyl-coenzyme A (CoA) for various biosynthetic purposes. The decarboxylation of pyruvate loses a carbon equivalent, and limits the theoretical carbon yield to only two moles of two-carbon (C2) metabolites per mole of hexose. This native route is a major source of carbon loss in biorefining and microbial carbon metabolism. Here we design and construct a non-oxidative, cyclic pathway that allows the production of stoichiometric amounts of C2 metabolites from hexose, pentose and triose phosphates without carbon loss. We tested this pathway, termed non-oxidative glycolysis (NOG), *in vitro* and *in vivo* in *Escherichia coli*. NOG enables complete carbon conservation in sugar catabolism to acetyl-CoA, and can be used in conjunction with CO₂ fixation¹ and other one-carbon (C1) assimilation pathways² to achieve a 100% carbon yield to desirable fuels and chemicals.

The word glycolysis was coined in the late 1800s when researchers noticed that the concentration of glucose decreased in yeast extracts³. It was only decades later that scientists understood that native glycolytic pathways oxidize glucose to form pyruvate, which is then converted to acetyl-CoA through decarboxylation for either further oxidation or biosynthesis of cell constituents and products, including fatty acids, amino acids, isoprenoids and alcohols. Various glycolytic pathways have been discovered, including the classic Embden–Meyerhof–Parnas (EMP) pathway, the Entner–Doudoroff (ED) pathway⁴, and their variations^{5,6}, which produce acetyl-CoA from sugars through oxidative decarboxylation of pyruvate mediated by the pyruvate dehydrogenase complex or through the anaerobic pyruvate-formate lyase reaction. Similarly, the Calvin–Benson–Bassham (CBB) and ribulose monophosphate (RuMP) pathways incorporate C1 compounds, such as CO₂ and methanol, to synthesize sugar phosphates, which then ultimately produce acetyl-CoA through pyruvate. Although the pyruvate route to acetyl-CoA, acetate⁷ and ethanol⁸ has been optimized for various purposes, the carbon loss problem has not been solved owing to inherent limitations in this pathway. Without using a CO₂ fixation pathway^{1,9,10}, the wasted CO₂ leads to a significant decrease in carbon yield. This loss of carbon has a major impact on the overall economy of biorefinery and the carbon efficiency of cell growth. On the other hand, re-fixing the lost CO₂ would incur energetic and kinetic costs.

Theoretically, it is possible to split sugars or sugar phosphates into stoichiometric amounts of acetyl-CoA in a carbon- and redox-neutral manner. Pathways without excess redox equivalents would be more efficient and could lead to maximal yields¹¹. However, no such pathway is known to exist. Here we constructed the cyclic NOG pathway (Fig. 1a) to break down sugars or sugar phosphates into the theoretical maximum amount of C2 metabolites without carbon loss.

The metabolic logic of the cyclic NOG pathway can be understood by breaking it down into three sections (Fig. 1a). First, fructose 6-phosphate (F6P) is the input molecule, and the pathway requires

an additional investment of two F6P molecules. Second, the three F6P molecules are broken down to three acetyl phosphate (AcP) and three erythrose 4-phosphate (E4P) molecules by the phosphoketolases. This irreversible step provides the first driving force for NOG. Third, these three E4P molecules then undergo carbon rearrangement to regenerate the two initially invested F6P molecules (Fig. 1b). The net reaction results in the irreversible formation of three AcP molecules. Phosphoketolases are known to have either F6P activity (termed Fpk) or xylulose 5-phosphate (X5P) activity (termed Xpk). Even though the product of Xpk is different from Fpk (glyceraldehyde 3-phosphate (G3P) is formed, as opposed to E4P), both are metabolically equivalent when Xpk is used in combination with transketolase (Tkt) (Supplementary Fig. 1a).

The regeneration of two F6P from three E4P can occur in several different ways. Figure 1b shows a fructose 1,6-bisphosphate (FBP)-dependent network, and Supplementary Fig. 1b shows a sedoheptulose 1,7-bisphosphate (SBP)-dependent network, with carbon rearrangement schemes illustrated by the colour-coded carbon atoms. The FBP-dependent network involves transaldolase (Tal), FBP aldolase (Fba) and fructose 1,6-bisphosphatase (Fbp). The SBP-dependent network does not involve Tal, but requires SBP aldolase and sedoheptulose 1,7-bisphosphatase. The two bisphosphatases provide the second irreversible driving force to ensure that carbon rearrangement proceeds towards the formation of F6P. Analogous systems of carbon rearrangement¹² are used in several natural pathways, such as the CBB, RuMP and the pentose phosphate pathways (see Supplementary Fig. 2).

Because there are two different possible phosphoketolase activities (Fpk and Xpk, shown in Supplementary Fig. 1a) and two variations of the carbon rearrangement networks (Supplementary Fig. 1bc), many combinations can be devised. For each of the carbon rearrangement networks, three configurations exist that form a basis^{13,14} to all other combinations (Fig. 2a–c): (1) NOG using only Fpk; (2) NOG using only Xpk; and (3) NOG using one Fpk with two Xpk activities but without the use of one type of Tkt reaction—the reversible conversion of F6P and G3P to E4P and ribose 5-phosphate (R5P). Figure 2a–c shows the three modes of NOG using the FBP-dependent carbon rearrangement network. In these configurations, the carbon rearrangement network and the phosphoketolase system are integrated such that the pathways appear different from those shown in Fig. 1b. Similarly, three modes of NOG can be derived for the SBP-dependent carbon rearrangement network. Combinations of these modes can generate infinite numbers of variations for NOG. NOG can be used in combination with other pathways, such as the CBB and RuMP pathways, which produce F6P from C1 compounds (Fig. 1a). In addition, NOG can also use pentose or triose sugar phosphates as input (Supplementary Fig. 3).

To validate the feasibility of this pathway experimentally, we constructed systems to demonstrate NOG both *in vitro* and *in vivo*. To construct an *in vitro* system, we cloned a putative phosphoketolase which has both Fpk and Xpk activities (F/Xpk) from *Bifidobacterium adolescentis* and Tal, Tkt, Fbp, ribulose-5-phosphate epimerase (Rpe),

¹Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, 420 Westwood Plaza, Los Angeles, California 90095, USA. ²Department of Bioengineering, University of California, Los Angeles, 420 Westwood Plaza, Los Angeles, California 90095, USA. ³Institute for Genomics and Proteomics, University of California, Los Angeles, 420 Westwood Plaza, Los Angeles, California 90095, USA.

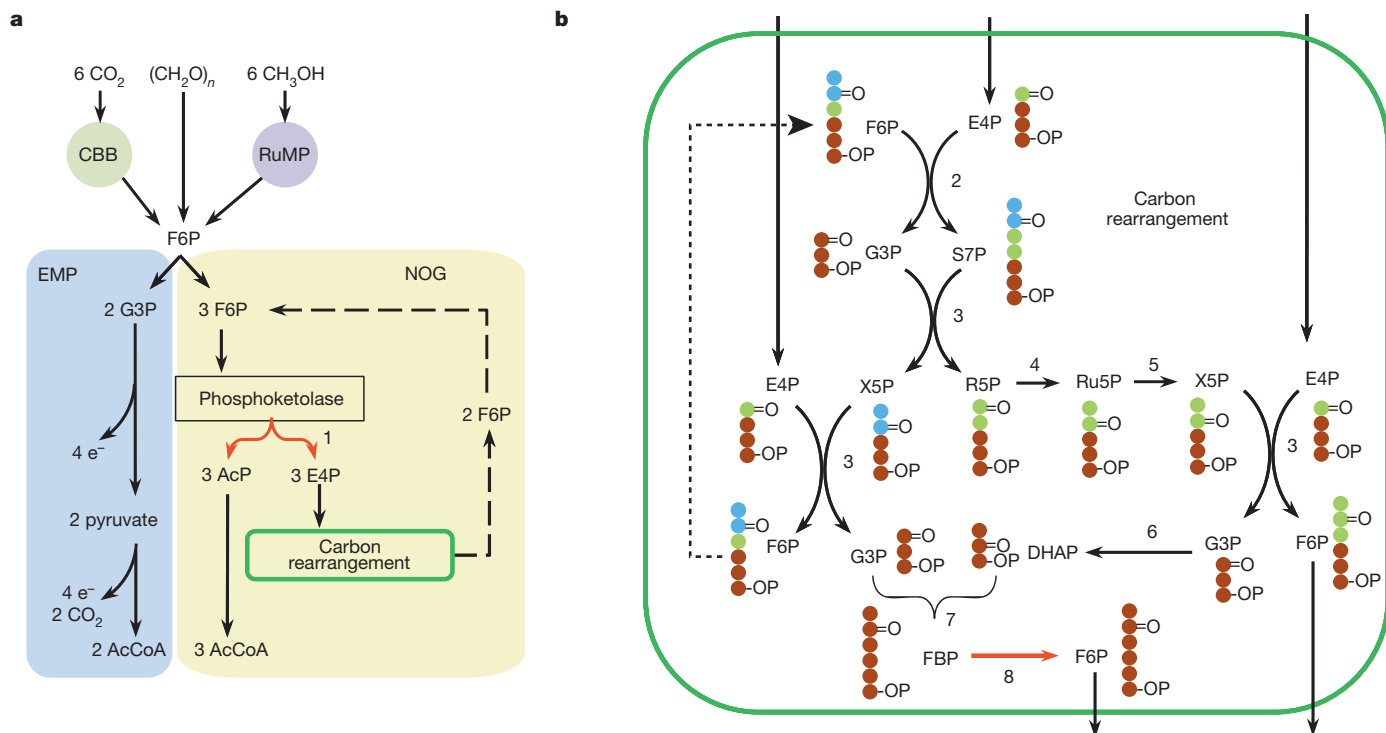


Figure 1 | Structure of oxidative (EMP) and non-oxidative glycolysis (NOG). **a**, Simplified schematic of EMP and NOG. **b**, An example of the carbon rearrangement network involving FBP. The red arrow indicates the irreversible phosphoketolase reaction. The carbon colour scheme illustrates

ribose-5-phosphate isomerase (Rpi) and acetate kinase (Ack) from *E. coli* with a His tag (Supplementary Figs 4 and 5) for one-step purification. Other enzymes, namely Fba, triose phosphate isomerase (Tpi), glycerol-3-phosphate dehydrogenase (Gpd), hexokinase (Glk), glucose-6-phosphate dehydrogenase (Zwf), phosphoglucose isomerase (Pgi) and phosphofructokinase (Pfk) were purchased. The His-tag enzymes were tested for activity (Supplementary Fig. 6 and Supplementary Table 1) and mixed together in a properly selected reaction buffer. This core system (F/Xpk, Tal, Tkt, Rpe, Rpi, Tpi, Fba, Fbp) was ATP and redox independent and consisted of eight core enzymes which convert one F6P molecule to three AcP molecules. The initial 10 mM F6P was completely converted to stoichiometric amounts of AcP (within error) at room temperature (25 °C) after 1.5 h (Fig. 3a). To extend the production

carbon rearrangement. Hydroxyl groups are not shown. Enzyme numbers are indicated: 1, phosphoketolase; 2, Tal; 3, Tkt; 4, Rpi; 5, Rpe; 6, Tpi; 7, Fba; 8, Fbp. DHAP, dihydroxyacetone phosphate; Ru5P, ribulose 5-phosphate.

further to acetate, Ack, phosphofructokinase (Pfk) and ADP were added to the *in vitro* NOG system. By adding a futile ATP-burning cycle^{15,16} using Pfk and Fbp, the complete conversion to acetate was possible (Fig. 3b). Similar *in vitro* NOG systems were tested on R5P and G3P, which produced nearly theoretical amounts of AcP at a ratio of 2.3 and 1.6, respectively (Fig. 3c). These *in vitro* results demonstrated the feasibility of NOG and paved the way for *in vivo* testing.

Next, NOG was engineered into the model organism *E. coli*. Xylose, instead of glucose, was used because it is the second most abundant sugar on Earth, and it avoids the use of the phosphotransferase system (PTS) for transport¹⁷, which is phosphoenolpyruvate (PEP) dependent and is associated with complex regulatory mechanisms¹⁸. To engineer NOG for xylose in *E. coli*, it was necessary to overexpress two enzymes:

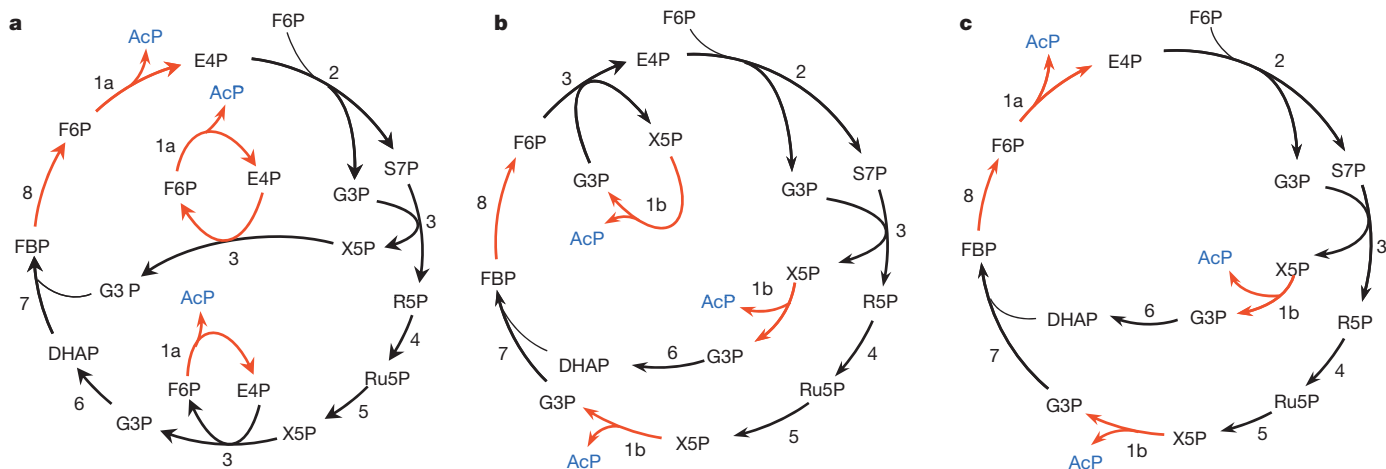


Figure 2 | Three FBP-dependent NOG networks. **a–c**, NOG using Fpk only (**a**), NOG using Xpk only (**b**) and NOG using F/Xpk (**c**). These configurations differ from those shown in Fig. 1 because the Xpk-linked Tkt has been

integrated with carbon rearrangement. The red arrows in **a–c** indicate irreversible reactions that drive the cycle. Enzyme numbers are defined in Fig. 1 legend, except: 1a, Fpk; 1b, Xpk.

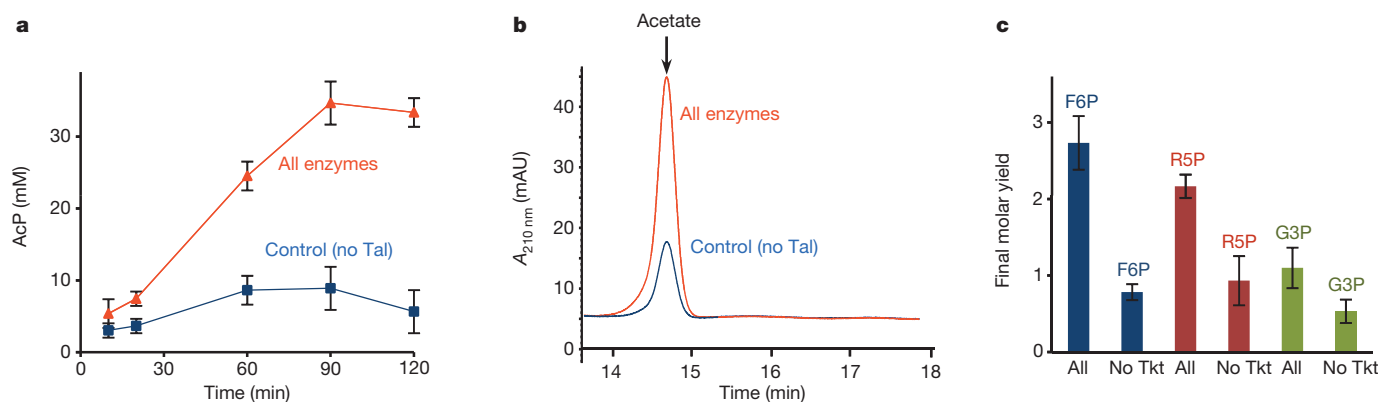


Figure 3 | In vitro NOG. **a**, *In vitro* conversion of F6P to AcP using eight purified core enzymes, including F/Xpk, Fbp, Fba, Tkt, Tal, Rpi, Rpe and Tpi. The starting F6P concentration was 10 mM. The red triangles are reactions with all eight enzymes present. The blue squares are reactions with all enzymes except Tal. **b**, *In vitro* conversion of F6P to acetate, determined by HPLC. The addition of Ack and Pfk allowed the complete conversion of AcP to acetate.

Acetate was monitored at 210 nm ($A_{210\text{nm}}$). **c**, Conversion of three sugar phosphates—F6P, R5P and G3P—to near stoichiometric amounts of AcP. 10 mM of each substrate was converted to AcP using the same core enzymes (denoted 'all'), whereas 'no Tkt' controls produced much less. *In vitro* enzyme assays were independently performed in triplicates and error bars indicate standard deviation (s.d.).

F/Xpk (encoded by *fxpk* from *B. adolescentis*) and Fbp (encoded by *E. coli fbp*). Other enzymes in NOG were natively expressed in *E. coli* under experimental conditions. The genes encoding F/Xpk and Fbp were cloned on a high-copy plasmid (pIB4) under the control of the P_{lacO_1} (ref. 19) isopropylthiogalactoside (IPTG)-inducible promoter. The plasmid was transformed into three *E. coli* strains: JCL16 (wild type), JCL166 (ref. 20) ($\Delta\text{ldhA } \Delta\text{adhE } \Delta\text{frdBC}$) and JCL118 ($\Delta\text{ldhA } \Delta\text{adhE } \Delta\text{frdBC } \Delta\text{pfkB}$). The latter two strains were used to avoid fermentative pathways that compete with NOG for F6P utilization (Fig. 4a). High-pressure liquid chromatography (HPLC) was used for monitoring xylose consumption and organic acid formation.

The dual expression of F/Xpk and Fbp in JCL188 was demonstrated by protein electrophoresis (Supplementary Fig. 5) and their activities were confirmed by a colorimetric enzyme assay (Fig. 4b). After an initial aerobic growth phase for cell growth and protein induction, high-cell-density cells were harvested and re-suspended in anaerobic minimal medium with xylose at a final optical density ($\text{OD}_{600\text{nm}}$) of 9. Anaerobic conditions were used to avoid the oxidation of acetate through the tricarboxylic acid (TCA) cycle. The wild-type host (JCL16) with plasmid pIB4 produced a mixture of lactate, formate, succinate and acetate from xylose, and the yield of acetate was quite low at about 0.4 acetates produced per xylose consumed. By introducing ΔldhA , ΔadhE and ΔfrdBC knockouts (JCL166), and with plasmid pIB4, the yield was increased to 1.1 acetates per xylose (5.1 g l^{-1} in 16 h) consumed. After further deleting *pfkB* (JCL118) and adding pIB4, the titre reached a maximum of only 3.6 g l^{-1} in 16 h. However, the yield reached the highest level of 2.2 acetates per xylose consumed, approaching the theoretical maximum of 2.5 moles of acetate per mole of xylose (Fig. 4c), and exceeding the theoretical maximum of 1.67 moles of acetate per mole of xylose supported by the EMP pathway only. Even though the rate of acetate production decreased, which is a complex result of enzymatic kinetics and regulation, the increase in yield supported the *in vivo* activity of NOG. Some succinate remained, presumably due to succinate dehydrogenase (*sdhABCD*) left over from the aerobic growth phase. These results indicate that NOG is also feasible *in vivo*. With further optimization, NOG represents a practical alternative for acetyl-CoA biosynthesis in various applications.

When NOG is used to degrade glucose, the net reaction is two ATP and three acetate molecules per glucose, while no reducing equivalents are generated. Therefore, NOG may not be important for normal cellular growth, which requires reducing equivalents and metabolites in the EMP pathway. These features may explain why NOG has not been found in nature. The challenges for re-routing glycolysis to NOG

in *E. coli* for growth involve (1) replacing the PTS-glucose transport system with a system that does not depend on PEP, (2) deleting the lower part of the EMP pathway, (3) adapting the *E. coli* strain to grow under this newly wired central metabolism, and (4) providing alternative source of reducing power. ^{13}C -labelled substrate can be used to determine the relative flux of each pathway based on the rearrangement pattern depicted in Fig. 1b. When a sugar (such as xylose or glucose) is used as the initial substrate to generate the input to NOG, acetate can be produced without carbon loss and without additional input of reducing equivalents. Once additional reducing power is provided in the form of hydrogen or formic acid²¹, NOG can be used to produce compounds that are more reduced than acetate, such as ethanol, 1-butanol, isoprenoids and fatty acids²².

The critical enzyme in NOG is F/Xpk, which is used in nature in the heterofermentative phosphoketolase pathway (PKP) and in the bifid shunt (Supplementary Fig. 7). PKP has a relatively low net ATP yield of one ATP per glucose consumed (Supplementary Table 2). Xpk has also been found in many organisms, such as *Clostridium acetobutylicum*, in which up to 40% of xylose is degraded by the PKP²³. *Bifidobacteria* utilize the unique bifid shunt²⁴ (Supplementary Fig. 7d), which oxidizes two glucoses into two lactates and three acetates. This fermentative pathway increases the ATP yield to 2.5 ATP per glucose, compared with 2 ATP per glucose for lactate fermentation through the EMP (Supplementary Table 2). In both phosphoketolase variants (PKP and the bifid shunt), G3P continues through the oxidative part of the EMP pathway to form pyruvate. Thus they are not able to directly convert glucose to three two-carbon compounds.

Acetogens, such as *Moorella thermoacetica*²⁵, achieve carbon conservation by fixing CO_2 emitted from pyruvate via the Wood–Ljungdahl pathway (Supplementary Fig. 7e). However, this pathway contains complex enzymes to overcome significant kinetic or thermodynamic barriers²⁶. By contrast, NOG contains no difficult enzymes and is amenable to heterologous expression (Supplementary Tables 3–5 and Supplementary Fig. 8). NOG can also be used in conjunction with C1 assimilation pathways that synthesize acetyl-CoA via pyruvate. When combined with the CBB cycle, NOG provides complete carbon conversion in the synthesis of acetyl-CoA from CBB intermediates such as F6P or G3P. NOG coupled with the CBB pathway requires less ATP to synthesize one molecule of acetyl-CoA. Perhaps more importantly than saving ATP is that NOG requires less turnover of Rubisco for acetyl-CoA synthesis (Supplementary Table 6). For each acetyl-CoA produced, only two CO_2 molecules need to be fixed by Rubisco, as opposed to three CO_2 in the case of the CBB–EMP pathway combination with pyruvate dehydrogenase. This improved efficiency would become

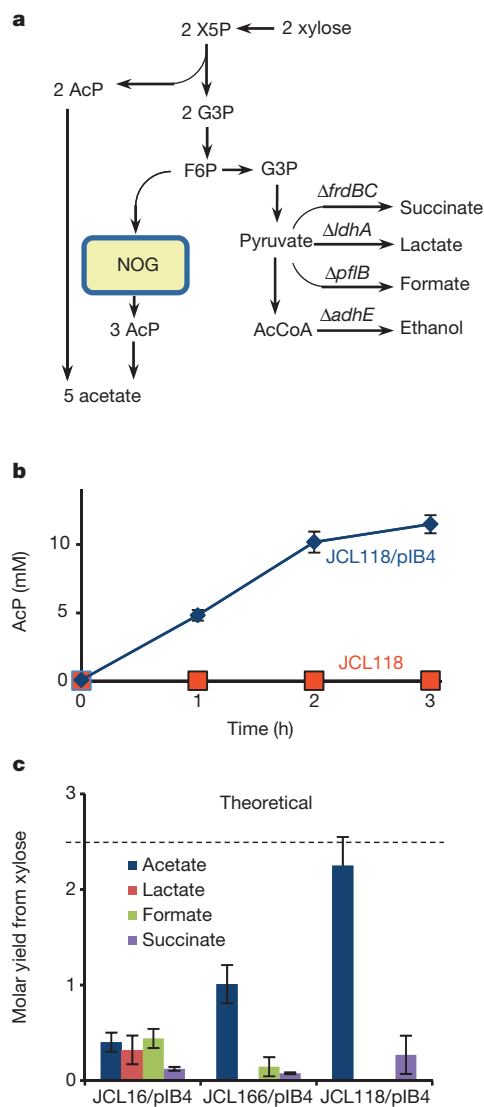


Figure 4 | In vivo conversion of xylose to acetate using NOG. **a**, Pathways in *E. coli* strains (JCL16, JCL166, JCL118) with NOG for converting xylose to acetate and other competing products (lactate, ethanol, succinate and formate production). Plasmid pIB4 was transformed into these strains for the expression of F/Xpk (from *B. adolescentis*) and Fbp (from *E. coli*) under the control of the P_{lacO_1} promoter. **b**, The expression of Fbp and F/Xpk in JCL118/pIB4 was tested by purifying the crude extract on a His-tag column, and then running a coupled colorimetric assay to test AcP formation. The control was JCL118 (without plasmid), which did not produce AcP. **c**, Xylose was converted to acetate and other products under anaerobic conditions. Strain JCL118 ($\Delta ldhA \Delta adhE \Delta frdBC \Delta pfIB$) produced a near theoretical ratio of acetate/xylose. *In vivo* production data were independently repeated three separate times from frozen glycerol stocks. Error bars indicate s.d.

especially important if an autotrophic organism was used for the biosynthesis of acetyl-CoA-derived products, such as 1-butanol²⁷ or fatty acids²⁸. Using NOG with the CBB pathway would represent a 50% increase in carbon efficiency in acetyl-CoA biosynthesis over native pathways. In view of the relatively low turnover number of Rubisco²⁹, this increased efficiency could allow faster production of acetyl-CoA-derived compounds. Given the variety of possible applications, NOG seems to be fundamentally important for carbon management.

METHODS SUMMARY

All NOG enzymes were constructed with amino-terminal His tags using the pQE9 from Qiagen. The genomic template for the *E. coli* genes was JCL16. JCL16 is wild-type *E. coli* strain BW25113 with F' transduced from XL-1 (Tet^R). JCL166 is JCL16 but with $\Delta ldhA$, $\Delta adhE$ and $\Delta frdBC$. JCL118 is similar

to JCL166 but with $\Delta pfIB$. The genomic template for F/Xpk was from *B. adolescentis* ATCC 15703. Many coupled enzyme assays measured the activity of the enzymes by measuring AcP at 505 nm or NAD(P)H at 340 nm. Plasmid pIB4 was constructed using the pZE12 (ref. 19) backbone with His-tagged Fbp and F/Xpk under the P_{lacO_1} promoter and was transformed into all production strains. For *in vivo* production, strains were grown aerobically on Luria-Bertani (LB) broth with 5% xylose, induced with 0.1 mM IPTG for 5 h, concentrated, and resuspended in fresh M9 5% xylose media. Organic acid and sugar concentration were monitored by HPLC using the HPX-87H column.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 15 August 2013.

Published online 29 September; corrected online 30 September and 30 October 2013 (see full-text HTML version for details).

- Bassham, J. *et al.* The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor. *J. Am. Chem. Soc.* **76**, 1760–1770 (1954).
- Yurimoto, H., Kato, N. & Sakai, Y. Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *Chem. Rev.* **5**, 367–375 (2005).
- Nelson, D. & Cox, M. *Lehninger Principles of Biochemistry* (W. H. Freeman, 2005).
- Peekhaus, N. & Conway, T. What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J. Bacteriol.* **180**, 3495–3502 (1998).
- Guterl, J.-K. *et al.* Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem* **5**, 2165–2172 (2012).
- Ye, X. *et al.* Synthetic metabolic engineering—a novel, simple technology for designing a chimeric metabolic pathway. *Microb. Cell Fact.* **11**, 120 (2012).
- Causey, T. & Ingram, L. Engineering the metabolism of *Escherichia coli* W3110 for the conversion of sugar to redox-neutral and oxidized products: homoacetate production. *Proc. Natl Acad. Sci. USA* **100**, 825–832 (2003).
- Tao, H., Gonzalez, R., Martinez, A. & Ingram, L. Engineering a homo-ethanol pathway in *Escherichia coli*: increased glycolytic flux and levels of expression of glycolytic genes during xylose fermentation. *J. Bacteriol.* **183**, 2979–2988 (2001).
- Berg, I. A., Kockelkorn, D., Buckel, W. & Fuchs, G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* **318**, 1782–1786 (2007).
- Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood–Ljungdahl pathway of CO₂ fixation. *Biochim. Biophys. Acta* **1784**, 1873–1898 (2008).
- Dugar, D. & Stephanopoulos, G. Relative potential of biosynthetic pathways for biofuels and bio-based products. *Nature Biotechnol.* **29**, 1074–1078 (2011).
- Meléndez-Hevia, E. & Isidoro, A. The game of the pentose phosphate cycle. *J. Theor. Biol.* **117**, 251–263 (1985).
- Liao, J. C., Hou, S.-Y. & Chao, Y.-P. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.* **52**, 129–140 (1996).
- Schuster, R. & Schuster, S. Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux. *Comput. Appl. Biosci.* **9**, 79–85 (1993).
- Daldal, F. & Fraenkel, D. Assessment of a futile cycle involving reconversion of fructose 6-phosphate to fructose 1, 6-bisphosphate during gluconeogenic growth of *Escherichia coli*. *J. Bacteriol.* **153**, 390–394 (1983).
- Patnaik, R., Roof, W. D., Young, R. F. & Liao, J. C. Stimulation of glucose catabolism in *Escherichia coli* by a potential futile cycle. *J. Bacteriol.* **174**, 7527–7532 (1992).
- Roseman, S. & Meadow, N. D. Signal transduction by the bacterial phosphotransferase system. *J. Biol. Chem.* **265**, 2993–2996 (1990).
- Gonzalez, R., Tao, H., Shanmugam, K. T., York, S. W. & Ingram, L. O. Global gene expression differences associated with changes in glycolytic flux and growth rate in *Escherichia coli* during the fermentation of glucose and xylose. *Biotechnol. Prog.* **18**, 6–20 (2002).
- Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I₁–I₂ regulatory elements. *Nucleic Acids Res.* **25**, 1203 (1997).
- Atsumi, S. *et al.* Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **10**, 305–311 (2008).
- Li, H. *et al.* Integrated electromicrobial conversion of CO₂ to higher alcohols. *Science* **335**, 1596 (2012).
- Gronenberg, L. S., Marcheschi, R. J. & Liao, J. C. Next generation biofuel engineering in prokaryotes. *Curr. Opin. Chem. Biol.* **17**, 462–471 (2013).
- Liu, L. *et al.* Phosphoketolase pathway for xylose catabolism in *Clostridium acetobutylicum* revealed by ¹³C-metabolic flux analysis. *J. Bacteriol.* **194**, 5413–5422 (2012).
- Yin, X., Chambers, J. R., Barlow, K., Park, A. S. & Wheatcroft, R. The gene encoding xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (*xfp*) is conserved among *Bifidobacterium* species within a more variable region of the genome and both are useful for strain identification. *FEMS Microbiol. Lett.* **246**, 251–257 (2005).
- Drake, H. L. & Daniel, S. L. Physiology of the thermophilic acetogen *Moraxella thermoacetica*. *Res. Microbiol.* **155**, 869–883 (2004).
- Ragsdale, S. W. Metals and their scaffolds to promote difficult enzymatic reactions. *Chem. Rev.* **106**, 3317–3337 (2006).
- Lan, E. I. & Liao, J. C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl Acad. Sci. USA* **109**, 6018–6023 (2012).

28. Liu, X., Sheng, J. & Curtiss, R. Fatty acid production in genetically modified cyanobacteria. *Proc. Natl Acad. Sci. USA* **108**, 6899–6904 (2011).
29. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose biphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl Acad. Sci. USA* **103**, 7246–7251 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was partially supported by National Science Foundation (NSF) grant MCB-1139318 and Department of Energy (DOE) grant

DEO-SC0006698. I.W.B. was supported by NSF Integrative Graduate Education and Research Traineeship (IGERT) grant no. 0903720.

Author Contributions I.W.B. and J.C.L. conceived the project, designed the experiments, analysed the data, and wrote the manuscript. I.W.B. performed experiments, and T.-S.L. assisted in experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.L. (jiaoj@ucla.edu).

METHODS

Culture media, chemicals, and conditions. *E. coli* cultures were grown in Luria-Bertani (LB) broth with appropriate antibiotics at 37 °C at 250 r.p.m. Antibiotics were added at the following concentrations: ampicillin, 100 µg ml⁻¹; and kanamycin, 50 µg ml⁻¹. All chemicals and substrates were purchased from Sigma-Aldrich unless otherwise indicated.

Plasmid construction. Plasmids for expressing N-terminal polyhexahistidine-tagged Tkt, Tal, Rpe, Rpi, Ack, Fbp and F/Xpk were constructed using a modification of the ligation-independent cloning (LIC)³⁰ method using pQE9 (Qiagen) as the vector backbone. The primers used are listed in Supplementary Table 6. The genomic template for the *E. coli* genes was JCL16 (BW25113 strain). The genomic template for F/Xpk was from *B. adolescentis* ATCC 15703. Plasmid pIB4 was constructed using pZE12 (ref. 19) as the vector backbone with His-tagged Fbp and F/Xpk under the control of the same P_LlacO₁ promoter.

Protein purification. All in-house His-tagged proteins were purified by affinity chromatography using His-Spin Protein Miniprep kit (Zymo Research). Protein concentration was measured using diluted samples with Coomassie Plus Assay Reagent (Pierce) and bovine serum albumin (Bio-Rad) as the standard curve.

Enzyme assays. All enzymes were assayed using the same enzyme buffer, consisting of 50 mM Tris pH 7.5, 5 mM MgCl₂, 5 mM potassium phosphate. Cofactors were added where needed. The enzymes Fba, Glk, Zwf, Tpi and Gpd were purchased from Sigma-Aldrich. Each individual enzyme assay is described later and the reaction schemes are illustrated in Supplementary Fig. 6. All substrates were purchased from Sigma-Aldrich.

Assay for Ack. Ack was measured using both enzyme-linked assay and colorimetric hydroxamate assay for AcP consumption. The enzyme-linked assay coupled ATP formation with the formation of NADPH with a molar extinction coefficient at 340 nm (ϵ_{340} equal to 6.2 mM⁻¹ cm⁻¹) using commercial Glk and Zwf. This assay included the addition of 1 mM ADP, 2 mM glucose, 0.2 mM NADP, 0.5 U Glk, 0.5 U Zwf and 5 mM AcP. The activities of the linked enzymes were not limiting. In the colorimetric assay, AcP is reacted to form the brown ferric acetyl-hydroxamate measured at 505 nm. This end-point assay was stopped by adding 40 µl of assay solution to 60 µl of 2 M hydroxylamine pH 6.5. After 10 min, the colouring reagent consisting of 40 µl of 15% trichloroacetic acid, 40 µl of 4 M HCl, and 40 µl of FeCl₃ in 0.1 M HCl was added. A standard curve using commercial lithium potassium AcP was used to relate absorbance to concentration, which showed a linear relationship between 0 to 15 mM in the conditions described. Thus, the hydroxamate method was also used to directly measure the consumption of AcP by Ack.

Assay for F/Xpk. F/Xpk activity was also assayed using both the enzyme-linked ultraviolet assay and the end-point colorimetric assay. For the ultraviolet method,

the assay was similar to Ack except for the addition of 1 mM of thiamine pyrophosphate (TPP), 2.5 µg of purified F/Xpk, and 10 mM of substrate. Fpk activity was measured using F6P as the substrate. To measure Xpk activity, the addition of more than 2 U of Rpe and Rpi each allowed the use of R5P as the substrate (X5P was not available). The hydroxamate method was also used to directly measure the production of AcP from F6P or R5P.

Assay for Tkt. His-tagged Tkt was assayed using an enzyme-linked system to NADH (ϵ_{340} = 6.2 mM⁻¹ cm⁻¹) consumption with Tpi and Gpd. Because X5P was unavailable commercially, excess Rpi and Rpe were added in order to use R5P as the initial substrate. 1 mM TPP, 0.2 mM NADH, 5 mM R5P and over 2 U Tpi, 0.5 U Gpd, over 2 U Rpe and over 2 U Rpi was added to the enzyme buffer.

Assay for Tal. Tal was assayed in the reverse direction and coupled to NADPH production by Zwf. 2 U Rpe, over 2 U Rpi, 5 mM R5P, 95 µg Tkt, 1.2 µg Tal, over 2 U Pgi and 0.5 U Zwf was added to the enzyme buffer.

Assay for Fbp. Fbp was also assayed by NADPH production. Over 2 U Pgi, 1 µg Fbp and 0.5 U Zwf was added to the enzyme buffer.

In vitro NOG to convert F6P, R5P, G3P to AcP. 10 mM of each substrate was used as a substrate in enzyme buffer with 1 mM TPP with an additional 50 mM potassium phosphate. The following eight purified enzymes were included in the reaction mix: 0.1 U F/Xpk, 0.5 U Tkt, 0.5 U Tal, over 2 U Tpi, over 2 U Rpe, over 2 U Rpi, 0.2 U Fbp and 0.5 U Fba. AcP concentration (by hydroxamate assay) was measured over 2 h.

Construction of in vivo NOG strains. JCL16 is BW25113 with F' transduced from XL-1 (Tet^R). JCL166 is JCL16 but with ΔdhA , $\Delta adhE$ and $\Delta frdBC$. JCL118 is the same as JCL166 but with $\Delta pflB$. Gene knockouts were made using respective donor strains from the Keio collection³¹ using P1 transduction.

In vivo NOG production from xylose. These strains were transformed with pIB4 and grown initially aerobically in LB broth with 5% xylose. At semi-log phase the cultures were induced for 5 h using 0.1 mM IPTG. The cells were then harvested, concentrated, and re-suspended anaerobically at an optical density (OD_{600 nm}) of 9 in minimal M9 media with 5% xylose. Products from xylose from JCL16 were compared to JCL166 and JCL118 with pIB4 and analysed by an Agilent 1200 HPLC. Organic acids (succinate, lactate, formate and acetate) were detected at 210 nm by the Aminex HPX-87H (Bio-Rad) column using isocratic 5 mM sulphuric acid as the mobile phase at 0.6 ml min⁻¹ flow rate with 20 µl injection volume at 30 °C column temperature. The refractive index detector was used to measure the concentration of xylose.

30. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074 (1990).

31. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

Discovery of new enzymes and metabolic pathways by using structure and genome context

Suwen Zhao^{1*}, Ritesh Kumar^{2*}, Ayano Sakai^{2*}, Matthew W. Vetting^{3*}, B. McKay Wood^{2*}, Shoshana Brown⁴, Jeffery B. Bonanno³, Brandon S. Hillerich³, Ronald D. Seidel³, Patricia C. Babbitt⁴, Steven C. Almo³, Jonathan V. Sweedler^{2,5}, John A. Gerlt^{2,5,6}, John E. Cronan^{2,6,7} & Matthew P. Jacobson¹

Assigning valid functions to proteins identified in genome projects is challenging: overprediction and database annotation errors are the principal concerns¹. We and others² are developing computation-guided strategies for functional discovery with ‘metabolite docking’ to experimentally derived³ or homology-based⁴ three-dimensional structures. Bacterial metabolic pathways often are encoded by ‘genome neighbourhoods’ (gene clusters and/or operons), which can provide important clues for functional assignment. We recently demonstrated the synergy of docking and pathway context by ‘predicting’ the intermediates in the glycolytic pathway in *Escherichia coli*. Metabolite docking to multiple binding proteins and enzymes in the same pathway increases the reliability of *in silico* predictions of substrate specificities because the pathway intermediates are structurally similar. Here we report that structure-guided approaches for predicting the substrate specificities of several enzymes encoded by a bacterial gene cluster allowed the correct prediction of the *in vitro* activity of a structurally characterized enzyme of unknown function (PDB 2PMQ), 2-epimerization of *trans*-4-hydroxy-L-proline betaine (tHyp-B) and *cis*-4-hydroxy-D-proline betaine (cHyp-B), and also the correct identification of the catabolic pathway in which Hyp-B 2-epimerase participates. The substrate-liganded pose predicted by virtual library screening (docking) was confirmed experimentally. The enzymatic activities in the predicted pathway were confirmed by *in vitro* assays and genetic analyses; the intermediates were identified by metabolomics; and repression of the genes encoding the pathway by high salt concentrations was established by transcriptomics, confirming the osmolyte role of tHyp-B. This study establishes the utility of structure-guided functional predictions to enable the discovery of new metabolic pathways.

We have applied a structure-guided strategy using metabolite docking to multiple proteins and enzymes in a metabolic pathway to discover a previously undocumented reaction, 4R-hydroxyproline betaine 2-epimerase (Hyp-B 2-epimerase; Fig. 1a), as well as the catabolic pathway by which tHyp-B is converted to α -ketoglutarate. The crystallographically determined unliganded structure of the uncharacterized ‘target’ as well as homology models for a binding protein and a second enzyme encoded by its genome neighbourhood were used to predict the Hyp-B 2-epimerase activity as well as those of downstream enzymes of the pathway.

The marine bacterium *Pelagibaca bermudensis* encodes an uncharacterized member of the enolase superfamily (National Center for Biotechnology Information GI number 114543141) in which two lysine residues of the TIM-barrel domain are positioned to function as acid-base catalysts^{6–8}. The New York SGX Research Consortium determined its structure (PDB 2PMQ) because it shared less than 30% sequence identity with structurally characterized enolase superfamily members. The only ligand was the Mg²⁺ that stabilizes the enolate anion intermediate obtained by abstraction of the α -proton of a carboxylate substrate.

The active site is sequestered from solvent by two closed loops and was therefore suitable for virtual metabolite docking for substrate prediction (Supplementary Fig. 1).

Figure 2 shows the genome neighbourhoods of the gene encoding 2PMQ (*hpbD*; Hyp-B 2-epimerase from its functional characterization (see below)) plus a putative *Paracoccus denitrificans* orthologue. The automated TrEMBL annotations (Supplementary Table 1) fail to assign the *in vitro* activity of HpbD or to identify the metabolic pathway. *P. bermudensis* is not genetically tractable; we therefore studied *P. denitrificans*, which encodes one HpbD orthologue and two sets of orthologues of most of the genes neighbouring the *P. bermudensis hpbD* gene. Genome neighbourhoods are ‘conserved’ for other putative orthologues (about 20 can be identified in the sequence databases at <http://sfld.rbvi.ucsf.edu/>).

The *in silico* ligand docking library (87,098 members) included the KEGG metabolite library⁹ as well as other potential enolase superfamily substrates such as dipeptides, *N*-acylated amino acids, acid sugars and the enolate anions obtained by abstraction of the α -proton (high-energy intermediates¹⁰) (Methods). The library was docked in the active site of HpbD by using Glide SP, and energy scoring functions rank-ordered the members of the library according to binding affinity. The best-scoring molecules were enriched with amino acid derivatives, especially proline analogues and *N*-capped amino acid derivatives (Fig. 3), permitting the prediction that HpbD is an amino acid racemase/epimerase, with the substrate probably having *N*-substitution.

The genome neighbourhood includes an ABC transporter with a periplasmic binding protein (HpbJ) annotated as binding ‘glycine betaine/L-proline’. The structure of a homologous binding protein with glycine betaine¹¹ (PDB 1R9L) was used as the homology model template (HpbJ and 1R9L share 48% sequence identity) (Methods). The binding site contains three tryptophan residues (Fig. 1b) that form a π -cation ‘cage’ for a quaternary ammonium (betaine), which may also be electrostatically stabilized by Glu 42, located 5.4 Å from the quaternary nitrogen. Thus, the homology model permitted the prediction that its ligand is a betaine. A library of 31 betaines was docked to the model; tHyp-B had the highest rank (Supplementary Table 2), so we predicted that HpbJ participates in the transport of tHyp-B. In addition, the HpbD active site contains Trp 320 and Asp 292, which are similarly positioned relative to the predicted binding pose of betaines (Fig. 1c). The structural basis for the predicted specificity of HpbJ therefore refined the prediction that the substrate for HpbD is a proline betaine, for example tHyp-B.

A homology model was constructed for the Rieske-type protein (HpbB1) using a homologue (PDB 3N0Q) as the template (60% sequence identity; Fig. 1d) (Methods). The active site resembled the binding sites in the betaine-binding proteins (aromatic residues and Glu 200); indeed, some Rieske-type proteins are betaine demethylases^{12–14} (Supplementary Fig. 2 and Supplementary Table 3). We therefore predicted that the

¹Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, USA. ²Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ³Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ⁴Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94143, USA. ⁵Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ⁶Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ⁷Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

*These authors contributed equally to this work.

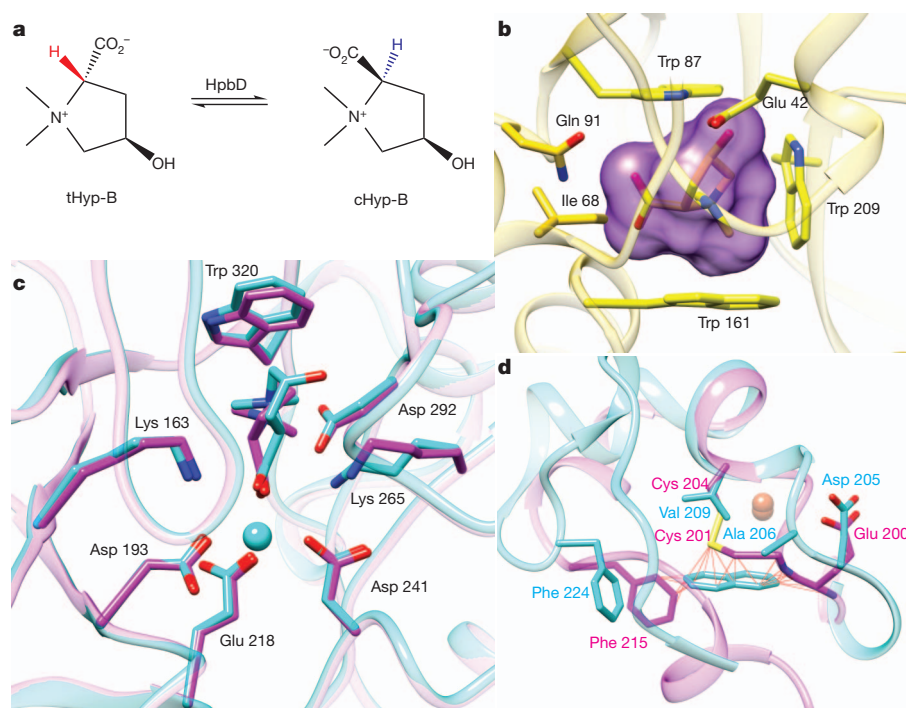


Figure 1 | Homology modelling and docking results for HpbD, HpbJ and HpbB1. **a**, The reaction catalysed by HpbD, the Hyp-B 2-epimerase. **b**, The binding site of the model of HpbJ, with the top-ranked ligand tHyp-B docked. The ligand surface is shown in magenta. **c**, Comparison of HpbD top-ranked docking pose of D-Pro-B (magenta) with the experimental pose of tHyp-B (cyan). The unliganded structure used in docking (PDB 2PMQ) and the subsequently determined liganded structure (PDB 4H2H) are shown in magenta and cyan, respectively. **d**, Superposition of the model of HpbB1 (magenta) and the closest characterized Rieske-type protein (cyan; PDB 1O7G, a naphthalene dioxygenase), showing that the active site of the model is too small to accept naphthalene as a substrate. Steric clashes identified by using a van der Waals overlap of 0.6 Å or more are shown in red lines.

substrate is a small betaine. (While this work was in progress, the X-ray structure of a Rieske-type Pro-B demethylase from *Sinorhizobium meliloti* (PDB 3VCP) was published¹⁵; its active site superimposed closely with our HpbB1 homology model.)

The results of library docking to the experimental *apo* structure of HpbD (2PMQ) and to homology models of HpbJ and HpbB1 enabled us to predict that HpbD uses Hyp-B or Pro-B as a substrate in a 1,1-proton transfer reaction. The betaines Gly-B and carnitine also were candidates (although these would be substrates for virtual reactions; the α -carbons are prochiral). Although more than 25 functions have been assigned to members of the enolase superfamily, including *N*-succinylamino acid racemases and dipeptide epimerases, no amino acid or amino acid betaine was known to be a substrate⁸.

tHyp-B, L-Pro-B, Gly-B and carnitine were incubated with HpbD in D₂O (Methods). The ¹H NMR spectra with tHyp-B, D-Pro-B and Gly-B revealed exchange of the α -proton with solvent deuterium (the latter being a virtual reaction); in addition, for tHyp-B, resonances associated with cHyp-B, the 2-epimer, were observed (Supplementary Fig. 3). These results are expected for a 1,1-proton transfer reaction that equilibrates

the configurations at carbon 2 of tHyp-B and Pro-B using two lysine acid–base catalysts.

The kinetic constants for tHyp-B and L-Pro-B were determined for both HpbD orthologues (Fig. 4b) (Methods). Although the k_{cat} values are large, the K_{m} values are also large, so the $k_{\text{cat}}/K_{\text{m}}$ values are modest. Betaines, including tHyp-B, are osmoprotectants accumulated by many bacteria, including pelagic (*P. bermudensis*) and plant-associated (*P. denitrificans*) species, to survive osmotic stress^{16–20}; their intracellular concentrations can approach molar levels^{21,22}. We determined that the intracellular concentration of Hyp-B is 170 mM in *P. denitrificans* grown on glucose in the presence of 0.5 M NaCl and 20 mM tHyp-B (Methods). Hyp-B 2-epimerase therefore probably functions with a high intracellular concentration of tHyp-B, so the kinetic constants are both physiologically reasonable and expected²³. That only four compounds were tested (tHyp-B, L-Pro-B, carnitine and Gly-B) and two have physiologically relevant kinetic constants confirms that pathway docking enables efficient functional prediction.

The 1.70-Å structure of HpbD was determined in the presence of tHyp-B (Methods, Supplementary Fig. 4 and Supplementary Table 4).

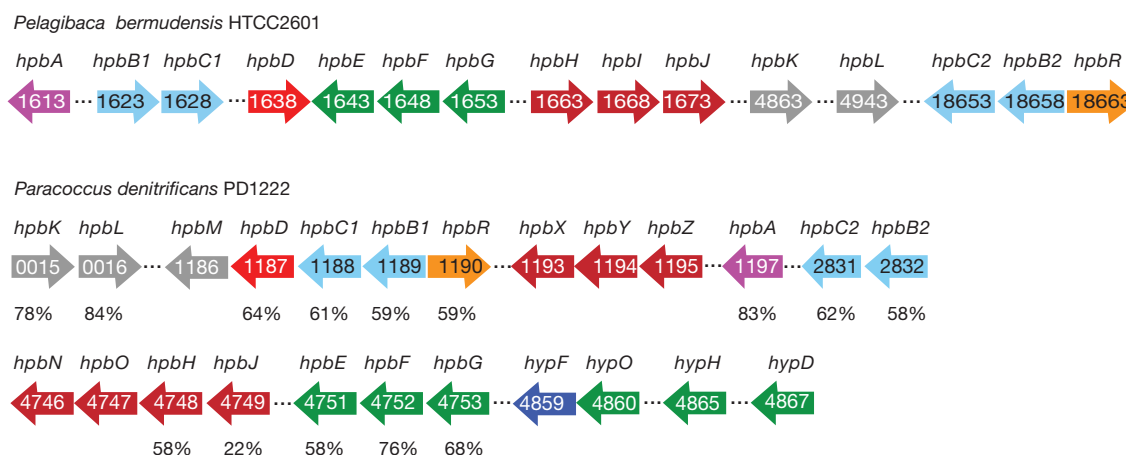


Figure 2 | Genome contexts of HpbD in *P. bermudensis* and the orthologous genes in *P. denitrificans*. The genes encoding orthologues are highlighted with the same colour; the sequence identities relating orthologues

in *P. bermudensis* and *P. denitrificans* are indicated. The ecological sources of tHyp-B would be seaweed (sargasso) for the Sargasso Sea bacterium *P. bermudensis*, and plants for the soil bacterium *P. denitrificans*.

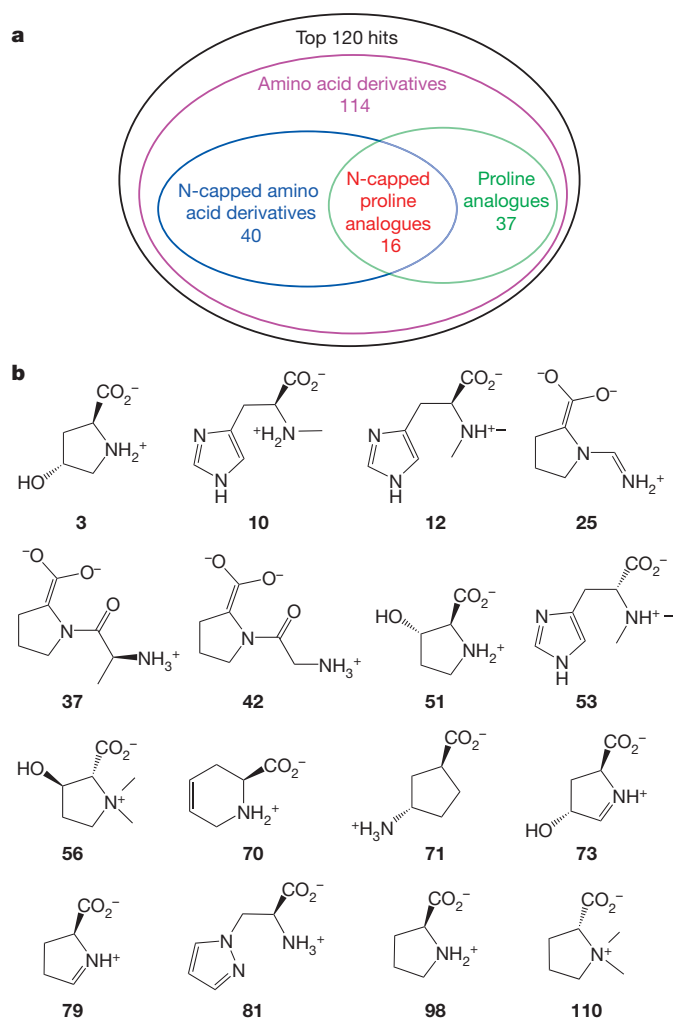


Figure 3 | Chemotype analysis of HpbD docking results. **a**, Enriched chemotypes in the top 120 hits. Most of them are amino acid derivatives, in which N-capped amino acid derivatives and proline analogues are the two most common subtypes. **b**, Proline analogues in the rank-ordered list of predicted ligands, illustrating the frequent occurrence of N-modified proline analogues. Pro-B, a substrate for HpbD, ranks at number 110 in the list (top 0.12% of the docking library).

The ligand electron density, with elevated *B*-factors, was interpreted as a mixture of tHyp-B (substrate) and cHyp-B (product) (Supplementary Fig. 5). The betaine forms a π -cation interaction with Trp 320 and is proximal to Asp 292, similar to the interactions in the Gly-B periplasmic binding protein (Fig. 1c). The predicted pose with D-Pro-B superimposes closely on the experimental pose, explaining the correct computation-based prediction of substrate specificity.

We also speculated that the binding proteins and enzymes encoded in the HpbD genome neighbourhoods constitute a catabolic pathway that degrades tHyp-B to α -ketoglutarate (Fig. 4a), with HpbD catalysing the first step in which tHyp-B is 2-epimerized to cHyp-B; that is, the *in vivo* activity of HpbD is Hyp-B 2-epimerase. Subsequently, HpbB1/HpbC1, the Rieske-type protein, catalyses the demethylation of cHyp-B to N-methyl cHyp; HpbA, a flavin-dependent enzyme, converts N-methyl cHyp to cHyp; HpbE, a D-amino acid oxidase, catalyses the oxidation of cHyp to its imino acid; HpbG, a member of the dihydrodipicolinate synthase superfamily²⁴, catalyses the dehydration of the 4-OH group and 'hydrolysis' of the 5-amino group to α -ketoglutarate semialdehyde; and HpbF, an aldehyde dehydrogenase, catalyses the oxidation of α -ketoglutarate semialdehyde to α -ketoglutarate. This pathway would permit the utilization of tHyp-B as a carbon and nitrogen source. The activities predicted for HpbE and HpbG were

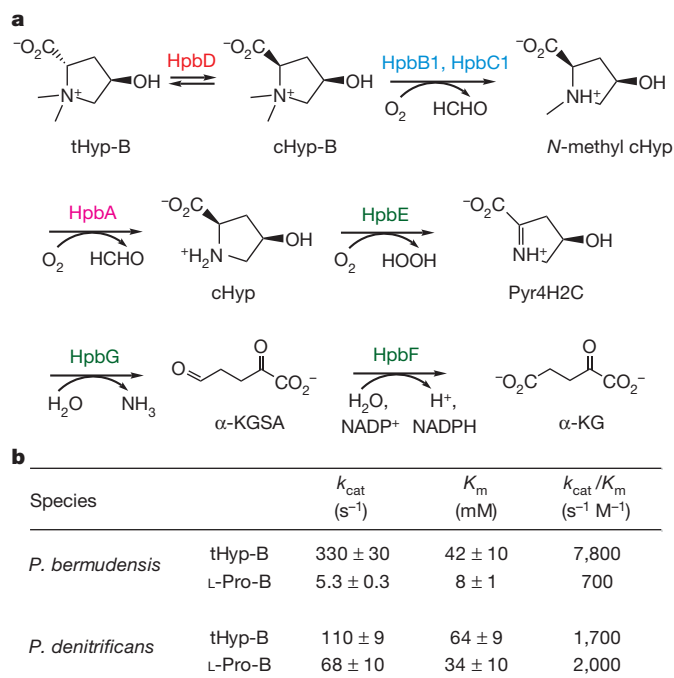


Figure 4 | Catabolic pathway for tHyp-B and kinetic constants for HpbD.

a, Catabolic pathway for tHyp-B. On the basis of the genome neighbourhood contexts in *P. bermudensis* and *P. denitrificans*, tHyp-B is epimerized to cHyp-B which undergoes two N-demethylation reactions to cHyp; cHyp is oxidized, dehydrated and deaminated, and finally oxidized to α -ketoglutarate (α -KG). Pyr4H2C, Δ^1 -pyrroline-4-hydroxy-2-carboxylate; α -KGSA, α -ketoglutarate semialdehyde. The enzymes are coloured as in Fig. 2. **b**, Kinetic constants for HpbD from *P. bermudensis* and its orthologue from *P. denitrificans*.

described recently in pathways for tHyp catabolism in *Pseudomonas aeruginosa*, *Pseudomonas putida* and *S. meliloti*^{25,26}; however, the sequences of HpbE and HpbG are so divergent (less than 35% sequence identity) that the cHyp oxidase and cHyp imino acid dehydratase/deaminase functions could not be assigned to the *P. bermudensis* and *P. denitrificans* enzymes without additional information (Supplementary Fig. 6).

When tHyp-B is used as an osmoprotectant (sea water is about 0.6 M NaCl), its catabolism should be depressed to maintain high intracellular concentrations. However, in the absence of osmotic stress, bacteria should be able to catabolize tHyp-B as a carbon and nitrogen source. *P. denitrificans* utilizes both tHyp-B and cHyp-B as carbon and nitrogen sources at low salt concentrations, as expected if the genome encodes the proposed catabolic pathway (Methods). Moreover, tHyp-B alleviates growth inhibition at high salt concentration (0.5 M NaCl) in glucose medium, arguing that *P. denitrificans* uses tHyp-B as an osmoprotectant (Supplementary Figs 7–9). Growth stimulation by tHyp-B in high-salt glucose medium could result from both osmoprotection and catabolism of tHyp-B. To address this possibility we used strain RPD4, which lacks the demethylases that convert the isomers of Hyp-B to the isomers of N-methyl Hyp and therefore cannot utilize tHyp-B or cHyp-B as a carbon source (Supplementary Table 8). Strain RPD4 grew almost as well on high-salt glucose medium in the presence of tHyp-B or cHyp-B as did the culture without salt supplementation, but growth on high-salt glucose medium in the absence of tHyp-B or cHyp-B was strongly inhibited (Supplementary Fig. 8). The results establish that tHyp-B and cHyp-B function as osmoprotectants.

We identified the metabolites obtained from tHyp-B at low salt concentrations (Methods). In addition to Hyp-B (21 mM; Supplementary Table 10), N-methyl Hyp and Hyp (the carbon-2 epimers cannot be distinguished), the predicted downstream Δ^1 -pyrroline-4-hydroxy-2-carboxylate, α -ketoglutarate semialdehyde and α -ketoglutarate were observed (Supplementary Figs 10 and 11). The metabolites were not detected with succinate as a carbon source. In high-salt glucose medium

containing tHyp-B, the intracellular concentration of Hyp-B was 170 mM (as expected for an osmolyte; Supplementary Table 10); however, its downstream metabolites were not detected. Thus, the flux through the pathway is regulated so that tHyp-B is not catabolized when it is needed as osmoprotectant^{19,20,27}. No Hyp-B was detected in cells grown on high-salt glucose medium, establishing that *P. denitrificans* lacks an anabolic pathway for tHyp-B.

We used quantitative PCR with reverse transcription (qRT-PCR) to investigate expression of the genes encoding the catabolic pathway (Methods and Supplementary Table 6). *P. denitrificans* encodes one orthologue of Hyp-B 2-epimerase (HpbD) and the FAD-dependent *N*-methyl Hyp demethylase but two orthologues of the remaining proteins and enzymes involved in the transport of tHyp-B and its catabolism (Fig. 2). The genes encoding the pathway are upregulated by tHyp-B and cHyp-B, as expected if their encoded proteins are involved in the catabolic pathway.

The effects of high salt concentration were determined using equimolar concentrations of glucose and either tHyp-B or cHyp-B. Salt (0.5 M NaCl) enhanced the expression of the transporters (HpbN/HpbO/HpbH/HpbJ and HpbX/HpbY/HpbZ). In contrast, salt decreased the expression of the genes encoding Hyp-B 2-epimerase (HpbD), both Hyp-B demethylases (HpbB1/HpbC1 and HpbB2/HpbC2) and the single *N*-methyl Hyp demethylase (HpbA) (Supplementary Table 6). Transport of tHyp-B/cHyp-B is required for uptake as osmolytes as well as carbon and nitrogen sources; expression of their transporters is enhanced, whereas epimerization and demethylation are suppressed, thereby allowing tHyp-B/cHyp-B to be retained as osmolytes.

The genes encoding the *P. denitrificans* pathway were individually disrupted by the insertion of antibiotic-resistance cassettes (Methods and Supplementary Table 7). The growth phenotypes are consistent with the predicted functions (Supplementary Discussion).

Here we have used homology modelling and metabolite docking to several proteins encoded by a gene cluster to guide the *in vitro* assignment of the previously undocumented Hyp-B 2-epimerase activity to 2PMQ, a structure determined by the Protein Structure Initiative. With knowledge of the catalytic capabilities of enzyme superfamilies, we also predicted the pathway that catabolizes cHyp-B to α -ketoglutarate. These predictions were verified by metabolomics and genetics. Finally, we used transcriptomics to demonstrate that Hyp-B 2-epimerase is a 'switch' that determines whether the tHyp-B is accumulated as an osmolyte or catabolized as carbon and nitrogen source.

Orthologues of HpbD can be identified in 20 microbial species (<http://sfld.rvbi.ucsf.edu/>), so both the *in vitro* activity and the *in vivo* functional assignments identified in this study can be extended to these proteins and organisms. Moreover, we expect that the Hyp-B 2-epimerase activity assigned to HpbD will be used to facilitate the discovery of the *in vitro* activities and *in vivo* functions of uncharacterized homologues in the enolase superfamily.

We propose pathway docking as an efficient strategy for predicting *in vitro* enzymatic activities and *in vivo* physiological functions. Additional refinements and applications of this strategy are in progress.

METHODS SUMMARY

The metabolite library was docked into the unliganded structure of HpbD (2PMQ) with program Glide SP followed by rescoring with MM-GBSA. Homology models for HpbJ and HpbB1 were made by PLOP v. 25.0; the betaine library was docked into the model of HpbJ with Glide XP. Kinetic constants for HpbD were measured by quantifying the change in optical rotation. HpbD was expressed with a carboxy-terminal hexahistidine tag; crystals were grown by sitting-drop vapour diffusion and flash-cooled in liquid nitrogen. Data were collected at the Advanced Photon Source beamline 31-ID (Lilly-CAT); the structure was determined by molecular replacement using the unliganded structure (2PMQ).

Metabolomics analyses by liquid chromatography Fourier transform mass spectrometry were performed with tHyp-B either as a sole source of carbon or as both an osmoregulant and the sole source of carbon²⁸. Gene expression profiles were studied with a Roche LightCycler 480. *P. denitrificans* gene disruption mutants were generated by conjugation²⁹ or electroporation³⁰.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 January; accepted 15 August 2013.

Published online 22 September 2013.

- Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLOS Comput. Biol.* **5**, e1000605 (2009).
- Gerlt, J. A. *et al.* The Enzyme Function Initiative. *Biochemistry* **50**, 9950–9962 (2011).
- Herrmann, J. C. *et al.* Structure-based activity prediction for an enzyme of unknown function. *Nature* **448**, 775–779 (2007).
- Song, L. *et al.* Prediction and assignment of function for a divergent *N*-succinyl amino acid racemase. *Nature Chem. Biol.* **3**, 486–491 (2007).
- Kalyanaraman, C. & Jacobson, M. P. Studying enzyme-substrate specificity *in silico*: a case study of the *Escherichia coli* glycolysis pathway. *Biochemistry* **49**, 4003–4005 (2010).
- Babbitt, P. C. *et al.* The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α -protons of carboxylic acids. *Biochemistry* **35**, 16489–16501 (1996).
- Gerlt, J. A., Babbitt, P. C. & Rayment, I. Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys.* **433**, 59–70 (2005).
- Gerlt, J. A., Babbitt, P. C., Jacobson, M. P. & Almo, S. C. Divergent evolution in enolase superfamily: strategies for assigning functions. *J. Biol. Chem.* **287**, 29–34 (2012).
- Tanabe, M. & Kanehisa, M. Using the KEGG database resource. *Curr. Protocols Bioinformatics* **38**, 1.12.1–1.12.43 (2012).
- Herrmann, J. C. *et al.* Predicting substrates by docking high-energy intermediates to enzyme structures. *J. Am. Chem. Soc.* **128**, 15882–15891 (2006).
- Schieffner, A. *et al.* Cation- π interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from *Escherichia coli*. *J. Biol. Chem.* **279**, 5588–5596 (2004).
- Goldmann, A. *et al.* Symbiotic plasmid genes essential to the catabolism of proline betaine, or stachydrine, are also required for efficient nodulation by *Rhizobium meliloti*. *FEMS Microbiol. Lett.* **115**, 305–311 (1994).
- Burnet, M. W. *et al.* The stachydrine catabolism region in *Sinorhizobium meliloti* encodes a multi-enzyme complex similar to the xenobiotic degrading systems in other bacteria. *Gene* **244**, 151–161 (2000).
- Wargo, M. J., Szwergold, B. S. & Hogan, D. A. Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. *J. Bacteriol.* **190**, 2690–2699 (2008).
- Daugherty, K. D. *et al.* Quaternary ammonium oxidative demethylation: X-ray crystallographic, resonance Raman, and UV-visible spectroscopic analysis of a Rieske-type demethylase. *J. Am. Chem. Soc.* **134**, 2823–2834 (2012).
- Larsen, P. I., Sydnnes, L. K., Landfald, B. & Strom, A. R. Osmoregulation in *Escherichia coli* by accumulation of organic osmolytes: betaines, glutamic acid, and trehalose. *Arch. Microbiol.* **147**, 1–7 (1987).
- Hanson, A. D. *et al.* Osmoprotective compounds in the Plumbaginaceae: a natural experiment in metabolic engineering of stress tolerance. *Proc. Natl Acad. Sci. USA* **91**, 306–310 (1994).
- Amin, U. S., Lash, T. D. & Wilkinson, B. J. Proline betaine is a highly effective osmoprotectant for *Staphylococcus aureus*. *Arch. Microbiol.* **163**, 138–142 (1995).
- Bernard, T., Pocard, J.-A., Berroul, B. & Le Rudulier, D. Variations in the response of salt-stressed *Rhizobium* strains to betaines. *Arch. Microbiol.* **143**, 359–364 (1986).
- Alloing, G., Travers, I., Sagot, B., Le Rudulier, D. & Dupont, L. Proline betaine accumulation and metabolism in alfalfa plants under sodium chloride stress. Exploring its compartmentalization in nodules. *J. Bacteriol.* **188**, 6308–6317 (2006).
- Burg, M. B., Kwon, E. D. & Kultz, D. Regulation of gene expression by hypertonicity. *Annu. Rev. Physiol.* **59**, 437–455 (1997).
- Kempf, B. & Bremer, E. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch. Microbiol.* **170**, 319–330 (1998).
- Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).
- Babbitt, P. C. & Gerlt, J. A. Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591–30594 (1997).
- Watanabe, S. *et al.* Identification and characterization of D-hydroxyproline dehydrogenase and Δ^1 -pyrroline-4-hydroxy-2-carboxylate deaminase involved in novel L-hydroxyproline metabolism of bacteria: metabolic convergent evolution. *J. Biol. Chem.* **287**, 32674–32688 (2012).
- White, C. E., Gavina, J. M., Morton, R., Britz-McKibbin, P. & Finan, T. M. Control of hydroxyproline catabolism in *Sinorhizobium meliloti*. *Mol. Microbiol.* **85**, 1133–1147 (2012).
- Gloux, K. & Le Rudulier, D. Transport and catabolism of proline betaine in salt-stressed *Rhizobium meliloti*. *Arch. Microbiol.* **151**, 143–148 (1989).
- Lenky, C. C., McEntyre, C. J. & Lever, M. Measurement of marine osmolytes in mammalian serum by liquid chromatography–tandem mass spectrometry. *Anal. Biochem.* **420**, 7–12 (2012).

29. Van Spanning, R. J. *et al.* A method for introduction of unmarked mutations in the genome of *Paracoccus denitrificans*: construction of strains with multiple mutations in the genes encoding periplasmic cytochromes C₅₅₀, C_{551i}, and C_{553i}. *J. Bacteriol.* **173**, 6962–6970 (1991).
30. Matsson, M., Ackrell, B. A., Cochran, B. & Hederstedt, L. Carboxin resistance in *Paracoccus denitrificans* conferred by a mutation in the membrane-anchor domain of succinate:quinone reductase. *Arch. Microbiol.* **170**, 27–37 (1998).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by cooperative agreements from the US National Institutes of Health (U54GM093342, U54GM074945 and U54GM094662). Molecular graphics and analyses were performed with the University of California, San Francisco (UCSF) Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at UCSF (supported by National Institutes of Health P41-GM103311). Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under contract no. DE-AC02-06CH11357. Use of the Lilly Research Laboratories Collaborative Access

Team (LRL-CAT) beamline at Sector 31 of the Advanced Photon Source was provided by Eli Lilly Company, which operates the facility.

Author Contributions S.Z., R.K., A.S., M.W.V., B.M.W., S.B., J.B.B., B.S.H., R.D.S., P.C.B., S.C.A., J.V.S., J.A.G., J.E.C. and M.P.J. designed the research. S.Z., R.K., A.S., M.W.V., B.M.W., J.B.B., B.S.H. and R.D.S. performed the research. S.Z., R.K., A.S., M.W.V., B.M.W., S.B., J.B.B., B.S.H., R.D.S., P.C.B., S.C.A., J.V.S., J.A.G., J.E.C. and M.P.J. analysed data. S.Z., R.K., A.S., M.W.V., B.M.W., S.B., J.B.B., B.S.H., R.D.S., P.C.B., S.C.A., J.V.S., J.A.G., J.E.C. and M.P.J. wrote the paper.

Author Information The atomic coordinates and structure factors for APO Hyp-B 2-epimerase (HpbD) and tHyp-B-liganded HpbD are deposited in the Protein Data Bank under accession numbers 2PMQ and 4H2H, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C.B. (babbitt@cgl.ucsf.edu), S.C.A. (steve.almo@einstein.yu.edu), J.V.S. (jsweedle@illinois.edu), J.A.G. (j-geritt@illinois.edu), J.E.C. (j-cronan@life.uiuc.edu) or M.P.J. (matt.jacobson@ucsf.edu).

METHODS

Homology modelling and docking. Sequence similarity network analysis. All sequences from the MLE subgroup in the Structure–Function Linkage Database (SFLD)⁷ were used in the MLE subgroup network analysis. BLAST analyses were performed with these sequences as queries in an all-by-all fashion. The details have been described previously⁴.

Sequences in the cHyp oxidase and Pyr4H2C deaminase networks were collected by BLAST, using red and blue dots in Supplementary Fig. 6 as queries, and 10^{-100} as BLAST E-value cutoff. The Pythoscape v. 1.0 program³¹ was used to make the two networks.

Homology modelling and docking. The models of HpbJ and HpbB1 were built with our in-house software Protein Local Optimization (PLOP, marketed as Prime by Schrödinger LLC). The template PDBs used for HpbJ and HpbB1 were 1R9L and 3N0Q, respectively. The sequence alignment of each pair of target and template was made by the L-INS-i method in MAFFT v. 6.925b (ref. 32). While constructing the models we included both the metal ions and the co-crystallized ligands (if any) from the templates. For docking, 2PMQ, the 1.72-Å X-ray apo structure of HpbD, was used. The structures were processed by Protein Preparation Wizard in Schrödinger Suite 2009 (ref. 33) before docking.

Two different libraries were used for docking in the active site of HpbD. The large metabolite library is the KEGG metabolite library plus potential substrates for members of the enolase superfamily not found in KEGG. The small library for focused docking to HpbD contained 31 betaines and betaine-like metabolites.

The KEGG metabolite library was generated by the following steps. First, we obtained 14,039 compounds from the KEGG COMPOUND database; then, we used LigPrep³⁴ in Schrödinger Suite 2009 to convert each compound from two dimensions to three dimensions and to enumerate up to 32 chiral forms. During this process, compounds with unspecified chemical groups (listed as 'R'), polymers and monatomic ions were automatically removed. Next, we removed compounds with molecular masses greater than 400 Da because we did not expect these to fit into the active site of HpbD, as well as duplicates generated by LigPrep preparation. We obtained 82,952 unique KEGG ligands.

Potential substrates for the enolase superfamily proteins include all dipeptides (formed by 20 standard amino acids), several types of N-capped (*N*-succinyl, *N*-acyl, *N*-formimino, *N*-formyl and *N*-carbamoyl) amino acids, acid sugars (monoacid sugars, diacid sugars, uronate sugars, 6-deoxy acid sugars and phospho sugars) and their corresponding enolates (that is, high-energy intermediates); these also were processed by LigPrep. After combining the KEGG metabolite library with these additional potential substrates for members of the enolase superfamily and removing duplicates, the library used for docking into the active site of HpbD contained 87,098 unique ligands.

The betaine library used for docking to the active site of HpbJ contains 31 betaines and betaine-like metabolites, including dimethylsulphoniopropionate (DMSP), ectoine, 5-hydroxyectoine and trigonelline; the compounds are listed in Supplementary Table 2. The members of this library also were processed by LigPrep.

Two docking methods were used. Glide SP docking followed by MM-GBSA was used with HpbD; the details have been described previously³⁵. The Glide XP docking method³⁶ was used with HpbJ.

In vitro activity measurements. Cloning, expression, and purification of the 2PMQ (HpbD). The protein sample was provided by the NYSGXRC structural genomics centre (PSI-2; U54GM074945).

Cloning, expression, and purification of the 2PMQ orthologue (HpbD) from *P. denitrificans*. The protein sample was provided by the NYSGXRC structural genomics centre.

Cloning, expression and purification of the HypF from *P. denitrificans*. The *hypF* gene was amplified by PCR using primers P17 and P18 and genomic DNA of *P. denitrificans* as a template. The PCR product was digested with *NdeI* and *BglII* and ligated to pET15b expression vector, yielding plasmid pRK9. The cloned HypF was expressed in *E. coli* BL21 (DE3) cells for protein purification. Luria–Bertani medium (4 l) was shaken at 20 °C and induced with 0.5 mM isopropyl β-D-thiogalactoside when the culture reached a D_{600} of 0.6. The cells were harvested after 24 h by centrifugation. The cells were resuspended in 100 ml of buffer containing 5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9, and 0.1 mM dithiothreitol (DTT). The suspension was lysed by sonication, and debris was cleared by centrifugation. The supernatant was applied to a Sepharose FF column charged with Ni^{2+} and eluted with a linear gradient (450 ml) of 60 mM to 1 M imidazole buffered with 0.5 M NaCl, 20 mM Tris-HCl pH 7.9, 0.1 mM DTT. The purest fractions were pooled and dialysed against 20 mM Tris-HCl pH 8.0, 0.1 mM DTT.

Screening HpbD activity by ^1H NMR. Epimerization/racemization of tHyp-B, L-Pro-B, Gly-B and carnitine were screened by disappearance of the α -proton in a D_2O -containing buffer by means of ^1H NMR. The reaction mixture contained 10 mM compound, 50 mM Tris-DCl pH 8.0, 10 mM MgCl_2 and 1 μM enzyme and

was incubated at 30 °C for 16 h before acquisition of the 500 MHz ^1H NMR spectrum.

Polarimetric assay for HpbD activity. Hyp-B 2-epimerase and L-Pro-B racemase activities were measured at 25 °C by quantifying the change in optical rotation. The assay was performed in a total volume of 0.8 ml in a cell with a path length of 100 mm, using a Jasco P-1010 polarimeter with a Hg 405-nm filter. Buffer conditions for the assay were 50 mM Tris-HCl pH 8.0 containing 10 mM MgCl_2 .

Polarimetric assay for HpbF activity. Hyp epimerase activity was measured at 25 °C by quantifying the change in optical rotation. The assay was performed in a total volume of 0.8 ml in a cell with a path length of 100 mm, using a Jasco P-1010 polarimeter with a Hg 405-nm filter. Buffer conditions for the assay were 50 mM sodium phosphate buffer pH 8.0 containing 1 mM DTT.

Structure determination. Expression of HpbD. Plasmid 9437a2BNT21p1, obtained from NYSGXRC stock clones³⁷, consists of a codon optimized HpbD gene in pSGX2, a derivative of pET26b (Novagen), with the amino-terminal methionine of HpbD changed to the sequence MAHHHHHHSL. The vector was transformed into Rosetta2 (DE3)pLysS competent cells (EMD Millipore) and plated on Luria–Bertani agar plates. Five to ten colonies were added to 75 ml of Luria–Bertani medium with 0.5% glucose and grown overnight at 37 °C. HpbD was expressed using 4 l of autoinduction medium at 25 °C (refs 38, 39). The starter culture and autoinduction medium were distributed equally among ten 2 l baffled flasks, and shaken at 300 r.p.m. for about 24 h to $D_{600} > 15$. All growth media contained 100 $\mu\text{g ml}^{-1}$ kanamycin and 50 $\mu\text{g ml}^{-1}$ chloramphenicol. Cells were pelleted and stored at –80 °C.

Purification of HpbD. All purification was performed at 4 °C. Cells were resuspended in $3\times$ (w/w) buffer A (50 mM HEPES pH 7.8, 150 mM NaCl, 20 mM imidazole, 10% (w/v) glycerol) supplemented with 0.1% (v/v) Tween 20 and disrupted by sonication. Cellular debris was removed by centrifugation, and the supernatant was applied to a 10-ml metal-affinity column (Ni^{2+} Sepharose High Performance; GE Healthcare) pre-equilibrated with buffer A. The column was washed with five column volumes of buffer A and subsequently eluted with two column volumes of the same buffer containing 300 mM imidazole. Eluted protein was pooled and applied to a 120-ml Superdex 200 column (GE Healthcare) equilibrated with buffer B (10 mM HEPES pH 7.5, 150 mM NaCl, 5% (v/v) glycerol). Fractions with more than 95% purity by SDS–PAGE analysis were pooled, concentrated by centrifugal ultrafiltration, snap-frozen in liquid nitrogen and stored at –80 °C.

Crystallization and structure solution of HpbD. Crystals were obtained by vapour diffusion at 18 °C using the sitting-drop vapour-diffusion method in 96-well IntelliPlates (Art Robbins). Equal volumes of protein (24.6 mg ml^{-1} in 10 mM HEPES pH 7.5, 150 mM NaCl, 5% (w/v) glycerol, 5 mM EDTA, 2 mM NiCl_2) and crystallization buffer (70% (v/v) 2-methyl-2,4-pentanediol, 0.1 M HEPES pH 7.5) were combined and equilibrated against 70 μl of crystallization buffer in the reservoir. Crystals grew as parallelograms measuring 0.05 mm \times 0.15 mm over a 1–2-week period. Crystals were soaked for 2 min in the reservoir solution supplemented with 200 mM *trans*-4-hydroxy-L-proline betaine (tHyp-B) and 50 mM MgCl_2 . Crystals were flash-cooled by immersion in liquid nitrogen, and subsequently stored and shipped to the Advanced Photon Source beamline 31-ID (Lilly-CAT). Data were collected at 100 K and a wavelength of 0.97929 Å. Crystals were rotated through 180° in 1° increments and the data were processed with MOSFLM⁴⁰ and scaled with SCALA⁴¹ in space group $P2_1$. The unliganded structure (APO) was determined by selenomethionine single-wavelength anomalous diffraction phasing by the NYSGXRC in 2007 from a carboxy-terminally hexahistidine-tagged protein (2PMQ; Supplementary Table 4), with one dimer per asymmetric unit. A single subunit from the unliganded structure was used as a search model in molecular replacement for the structural determination of the liganded structure. PHASER⁴² within the refinement package PHENIX⁴³ located eight subunits, which could subsequently be assembled into the molecular octamer. Several rounds of manual rebuilding and ligand and water fitting within the molecular graphics program COOT⁴⁴ followed by refinement in PHENIX were performed to finalize the structure. Several iodine atoms (seven or eight per subunit), originating from the synthesis of the substrate, were modelled into difference density peaks with features suggestive of bound iodine. The geometry restraints for tHyp-B were produced with the PRODRG2 server⁴⁵. The density was fitted equally well by tHyp-B and the product cHyp-B. It is presumed that the protein is active in the crystalline form and that the density is most probably a mix of substrate and product; however, only tHyp-B was used in refinement. The final structure has 98.6% of its residues in favoured regions of the Ramachandran plot, and 0.0% in disallowed regions (4H2H; Supplementary Table 4). The liganded structure (4H2H) superimposes with the APO structure (subunit A on subunit A) with a root mean squared deviation of 0.25 Å over 366 aligned C α atoms with no substantial changes to the structure on ligand binding.

Microbiology. Bacterial strains and growth conditions. *P. denitrificans* PD1222 wild-type and mutant strains were grown in minimal medium containing (in grams per litre) K_2HPO_4 6.0, KH_2PO_4 4.0, sodium molybdate 0.15, $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ 0.2,

CaCl₂ 0.04, MnSO₄·2H₂O 0.001, FeSO₄·7H₂O 1.1, with or without 1.6 g of NH₄Cl as nitrogen source. *P. denitrificans* PD1222 was grown aerobically at 30 °C in minimal medium supplemented with either glucose/succinate, tHyp-B or methanol at the same concentration (20 mM). *E. coli* strain TOP10 (Invitrogen) was used for plasmid maintenance, propagation and cloning purposes. *E. coli* strain S17-1 was used for conjugation⁴⁶. Strains used are listed in Supplementary Table 7. To study the role of tHyp-B in osmoprotection, cultures were grown in minimal medium with glucose, in the presence or absence of 500 mM NaCl. *E. coli* cultures were grown at 37 °C in Luria–Bertani medium. Antibiotics were used at the following concentrations (in µg ml⁻¹): kanamycin sulphate 50, chloramphenicol 35, sodium ampicillin 100.

Construction of disruption mutants. Gene inactivation mutant strains were generated in *P. denitrificans* by conjugation²⁹ or electroporation³⁰.

Molecular biology protocols. Chromosomal DNA was isolated from 3–5 ml of *P. denitrificans* PD1222 cell cultures with a DNeasy Blood and Tissue Kit (Qiagen) or a Wizard Genomic DNA Purification Kit (Promega). Restriction enzymes, DNA polymerases and T4 DNA ligases were purchased from New England Biolabs, Fermentas, Invitrogen or Promega. Plasmids were prepared from *E. coli* TOP10 cells with a Plasmid Mini Kit (Qiagen).

Gene disruption plasmids. Most plasmids for the construction of gene disruptions were obtained by a standard protocol in which the appropriate chromosomal segments were amplified from *P. denitrificans* PD1222 genomic DNA using *Pfu* polymerase followed by insertion of the PCR products into the pGEM T Easy vector (Promega). The resulting plasmids were then digested with *EcoRV* (pRK1), *NruI* (pRK2), *BmgBI* (pRK4) and ligated to a 900-base-pair (bp) fragment blunt-ended chloramphenicol resistance (*cat*) cassette. These plasmids were then used as PCR templates with the same primers and the products were ligated to vector pSUP202, which had been digested with *EcoRI* and treated with the Klenow fragment of DNA polymerase I, plus the four dNTPs to give the plasmids used for gene disruption. Primer sets P1 + P2, P3 + P4 and P7 + P8 gave rise to plasmids pRK1, pRK2 and pRK4, respectively. Plasmid pRK5 was obtained similarly by using primers P9 and P10 except that the original PCR product was inserted into vector pCR2.1-TOPO (Invitrogen) and the *cat* cassette was inserted into the *HincII* site. Plasmid pRK3 was obtained similarly by using primers P5 and P6 except that a 1,400-bp kanamycin resistance cassette was inserted into the *BmgBI* site of the intermediate pGEM T Easy construct. For plasmids pRK6 and pRK8, the products of PCR (primers P11 + P12 and P15 + P16, respectively) amplification from chromosomal DNA were inserted into pETBlue-1 (Novagen) and the kanamycin resistance cassette was inserted into the *SfoI* site of this plasmid. Plasmid pRK7 was obtained by the same manipulations with primers 13 and 14 except that the chloramphenicol cassette was inserted into the *SfoI* site.

Expression plasmids. These plasmids were constructed as above except that the PCR products obtained from chromosomal DNA contained the promoter and coding sequences and were directly ligated to vector pSUP404.2 that had been digested with *EcoRI* and treated with the Klenow fragment of DNA polymerase I plus the four dNTPs. Primers P19 + P2, P20 + P4 and P21 + P10 gave rise to plasmids pRK10, pRK11 and pRK12, respectively.

High-level protein expression plasmids pRK9 and pRK13 were obtained by insertion of the PCR products obtained from chromosomal DNA into pGEM T Easy. Plasmid pRK9 was obtained by ligation of the *NdeI* and *BglII* *hypF* fragment of the intermediate plasmid to pET15b digested with *NdeI* and *BamHI*, whereas pRK13 resulted from ligation of the *hypO* *NdeI*–*BpI* fragment of the pGEM T Easy intermediate plasmid into pET15b cut with the same enzymes.

Cell preparation for gene expression analysis. *P. denitrificans* PD1222 wild-type or mutant cultures were grown in 10 ml of minimal medium with 20 mM glucose as carbon source to a *D*₆₀₀ of 0.4. The cells were pelleted by centrifugation (4,000g for 10 min at 4 °C). The cell pellet was washed twice and resuspended in 10 ml of minimal medium lacking a carbon source. The cultures were divided into two 5-ml aliquots. Glucose or succinate was added to one aliquot, and Hyp-B (either *trans* or *cis*) or methanol was added to the other as carbon source followed by aerobic growth at 30 °C for 1 h before cell harvesting. For studying the effect of salt stress on gene expression, NaCl was added to the minimal medium to 500 mM; glucose or succinate was the carbon source, and Hyp-B was the osmoprotectant.

RNA sample preparation. For preparation of RNA samples, 1.0–5.0 ml of an actively growing *P. denitrificans* PD1222 culture (*D*₆₀₀ 0.5–0.6) was added to two volumes of RNeasy Protect Bacteria Reagent (Qiagen). After vortex-mixing for 10 s the solution was incubated for 5 min at 22 °C. The cells were pelleted by centrifugation (10,000g for 5 min at 4 °C), the supernatant was decanted and the remaining liquid was removed. RNA isolation was performed on ice in an RNeasy-free environment using an RNeasy Mini Kit (Qiagen), following the protocols for bacteria given by Qiagen. RNA concentrations were determined by absorption at 260 nm, with one absorbance unit corresponding to 44 µg ml⁻¹ RNA. Isolated RNA was analysed by agarose gel electrophoresis, and spectrophotometrically in

the Nanodrop (Thermo) using the ratios *A*₂₆₀/*A*₂₈₀ and *A*₂₆₀/*A*₂₃₀; the *A*₃₅₀/*A*₂₂₀ absorption spectra were used to assess sample integrity and purity. RNA preparation purity and absence of DNA was validated by agarose gel electrophoresis and control PCR reactions. The RNA preparations were stored at –80 °C until use.

qRT-PCR. Reverse transcription was performed on 1 µg of total RNA by using the RevertAid H Minus First Strand complementary DNA synthesis kit (Fermentas) in accordance with the manufacturer's protocol. Of the cDNA, 1 µl was used in separate PCR reactions of 20 µl for each gene. Minus-RT controls were performed to test for genomic DNA contamination in each RNA sample. Primers were designed by the Universal probe library system (Roche Applied Science). The primer length was 21–27 nucleotides, with a theoretical *T*_m of 58–60 °C. The amplicon size ranged from 66 to 110 bp. Real-time PCR was performed in 96-well plates with a Roche LightCycler 480. The 200-µl PCR mix was prepared by adding to 1 µl cDNA template 2 µl of forward and reverse primers (final concentration 150 nM of each primer) and 10 µl of SYBR 2× Concentration Green Master Mix (Roche). The PCR conditions were: one cycle at 95 °C for 5 min, 40 cycles of amplification at 95 °C for 15 s, 60 °C for 1 min. Finally, a dissociation program was applied with one cycle at 95 °C for 15 s, 60 °C for 1 min and 95 °C for 15 s. The efficiency of all the primers used in qRT-PCR was calculated as 97 ± 2%. The gene expression data were expressed as crossing point (CP) values. The 16S rRNA was used as a reference gene. The data was analysed by the 2^{-ΔΔC_T} (Livak) method⁴⁷. Data presented are the average of five biological replicates. Primer sequences are provided in Supplementary Table 9.

Metabolomics. Metabolomics analysis of the *P. denitrificans* tHPB degradation pathway. Liquid chromatography–Fourier transform mass spectrometry metabolomics of whole-cell extracts were performed with samples of *P. denitrificans* fed with tHyp-B with or without osmotic stress. First, these experiments were performed as a time course in minimal medium with or without the addition of tHyp-B. Succinate minimal medium grown cells were diluted 1:500 into 250 ml of minimal medium with tHyp-B as the sole carbon source and grown to a *D*₆₀₀ of about 0.7 (about 18 h). The culture was harvested by centrifugation (4,000g for 10 min at 4 °C), washed, and resuspended in minimal medium without carbon source. The cell suspension was then depleted of catabolic metabolites by incubation for 15 min at 30 °C before being transferred to ice. After verification of the concentrated cell density by multiple 50-fold dilutions into minimal medium (calculated *D*₆₀₀ 26.7 ± 0.2), 1-ml aliquots of cell suspensions with a *D*₆₀₀ of 6 were prepared in Eppendorf tubes on ice in minimal medium. tHyp-B (20 mM) was added quickly to half of the samples before incubation in a 30 °C water bath; at time points of 0, 2, 5 and 15 min, samples were transferred on ice to a centrifuge at 4 °C for pelleting at 16,000g for 2 min. The supernatant was then removed and the cell pellets were flash-frozen in liquid nitrogen. The process of collecting time-point samples from 30 °C to liquid nitrogen took about 3 min. Samples were stored at –80 °C before analysis.

To ascertain the effect of osmotic stress on the levels of the metabolites from tHyp-B catabolism, *P. denitrificans* was grown as described above except that a replicate culture with 0.5 M NaCl was also prepared. After these cultures reached a *D*₆₀₀ of about 0.5, each was concentrated by centrifugation, aliquoted into 1 ml of cells with a *D*₆₀₀ of 6, pelleted by centrifugation, separated from the supernatant, and frozen for analysis.

Metabolomics analysis of these samples followed the procedure from ref. 48. The cell pellets were extracted with 0.375 ml of 10 mM ammonium bicarbonate (pH 9.2) in 90% acetonitrile by pipetting followed by 15 min of vortex-mixing at 22 °C. The extraction was cleared of cell debris by two rounds of centrifugation at 16,000g before analysis. Samples were applied to a custom 11-T liquid-trap quadrupole Fourier transform mass spectrometer (Thermo-Fisher Scientific) with an Agilent 1200 high-performance liquid chromatography (HPLC) system equipped with a Sequant Zic-HILIC column (2.1 mm × 150 mm) previously equilibrated with the extraction buffer (solvent B). Solvent A was 10 mM ammonium bicarbonate pH 9.2. Each extracted sample was injected in 100 µl for three separate chromatographic runs. The samples were eluted at a flow rate of 200 µl min⁻¹ with the following elution profile: 100% B for 17 min, a linear gradient from 100% B to 40% B over 3 min, and another linear gradient from 40% to 100% B over 15 min. Data were collected at a resolution of 50,000 with full scan set to *m/z* 100–1,000, and duplicate samples were analysed individually in either positive or negative mode.

Data analysis was performed with the Qualbrowser application of Xcalibur (Thermo-Fisher Scientific) (Supplementary Figs 10 and 11). For metabolites, tHyp-B, *N*-methyl cHyp, cHyp and α-ketoglutarate, standards were run to verify retention time. Unfortunately, the *P. denitrificans* samples seemed to damage the Zic-HILIC column over time, because after dozens of runs many of the peaks broadened and showed longer retention times. Although HPLC analyses of betaine derivatives of biological origin have been reported^{48,49–51}, the metabolic analysis of bacteria with the use of betaines has not. It is assumed that the extremely large concentrations of tHyp-B accumulated in *P. denitrificans* cells were to blame as a result of overloading the column.

31. Barber, A. E. & Babbitt, P. C. Pythoscape: a framework for generation of large protein similarity networks. *Bioinformatics* **28**, 2845–2846 (2012).
32. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
33. Suite, S. 2009 *Protein Preparation Wizard*; *Epik version 2.0*; *Impact version 5.5*; *Prime version 2.1* (Schrödinger LLC, 2009).
34. Suite, S. 2009 *LigPrep, version 2.3* (Schrödinger LLC, 2009).
35. Kalyanaraman, C., Bernacki, K. & Jacobson, M. P. Virtual screening against highly charged active sites: identifying substrates of α - β barrel enzymes. *Biochemistry* **44**, 2059–2071 (2005).
36. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* **49**, 6177–6196 10.1021/jm051256o (2006).
37. Sauder, M. J. *et al.* High throughput protein production and crystallization at NYSGXRC. *Methods Mol. Biol.* **426**, 561–575 (2008).
38. Fox, B. G. & Blommel, P. G. Autoinduction of protein expression. *Curr. Protocols Protein Sci.* 5.23.1–5.23.18 (2009).
39. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
40. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 48–57 (2006).
41. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
42. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
43. Zwart, P. H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 (2008).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **D60**, 2126–2132 (2004).
45. Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
46. Simon, R., Priefer, U. & Puhler, A. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Nature Biotechnol.* **1**, 784–791 (1983).
47. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).
48. Erb, T. J. *et al.* A RubisCO-like protein links SAM metabolism with isoprenoid biosynthesis. *Nature Chem. Biol.* **8**, 926–932 (2012).
49. Chambers, S. T. & Kunin, C. M. Isolation of glycine betaine and proline betaine from human urine. Assessment of their role as osmoprotective agents for bacteria and the kidney. *J. Clin. Invest.* **79**, 731–737 (1987).
50. Oufir, M. *et al.* Simultaneous measurement of proline and related compounds in oak leaves by high-performance ligand-exchange chromatography and electrospray ionization mass spectrometry for environmental stress studies. *J. Chromatogr. A* **1216**, 1094–1099 (2009).
51. Li, C., Hill, R. W. & Jones, A. D. Determination of betaine metabolites and dimethylsulfoniopropionate in coral tissues using liquid chromatography–time-of-flight mass spectrometry and stable isotope-labeled internal standards. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **878**, 1809–1816 (2010).

Meiotic chromosome structures constrain and respond to designation of crossover sites

Diana E. Libuda¹, Satoru Uzawa², Barbara J. Meyer² & Anne M. Villeneuve^{1,3}

Crossover recombination events between homologous chromosomes are required to form chiasmata, temporary connections between homologues that ensure their proper segregation at meiosis¹. Despite this requirement for crossovers and an excess of the double-strand DNA breaks that are the initiating events for meiotic recombination, most organisms make very few crossovers per chromosome pair². Moreover, crossovers tend to inhibit the formation of other crossovers nearby on the same chromosome pair, a poorly understood phenomenon known as crossover interference^{3,4}. Here we show that the synaptonemal complex, a meiosis-specific structure that assembles between aligned homologous chromosomes, both constrains and is altered by crossover recombination events. Using a cytological marker of crossover sites in *Caenorhabditis elegans*⁵, we show that partial depletion of the synaptonemal complex central region proteins attenuates crossover interference, increasing crossovers and reducing the effective distance over which interference operates, indicating that synaptonemal complex proteins limit crossovers. Moreover, we show that crossovers are associated with a local 0.4–0.5-micrometre increase in chromosome axis length. We propose that meiotic crossover regulation operates as a self-limiting system in which meiotic chromosome structures establish an environment that promotes crossover formation, which in turn alters chromosome structure to inhibit other crossovers at additional sites.

Although crossovers mature in the context of assembled synaptonemal complexes (SCs), whether the SC functions in crossover interference has long been debated^{6–8}. Largely on the basis of analysis in budding yeast, it has been argued for the last decade that the SC is irrelevant for interference⁶. Consequently, suggestive evidence that the SC might be involved in interference in other organisms has gained little traction^{9,10}. The debate persists in part because of inherent limitations of some assays used to score crossovers (see Supplementary Discussion). We overcome this problem by taking advantage of the fact that the *C. elegans* crossover-promoting protein COSA-1, which forms foci at nascent crossover sites *in vivo*, serves as a robust cytological marker of the crossovers that elicit and respond to crossover interference⁵.

C. elegans exhibits robust crossover interference, with chromosome pairs normally undergoing only a single crossover². Using COSA-1 foci, we assessed whether the SC central region proteins (the SYP proteins) function in limiting the number of crossovers per chromosome pair. SYPs are essential for crossover formation^{11–13}, precluding the use of *syp* null mutants for this analysis. To circumvent this issue, we used RNA interference (RNAi) to partially deplete SYP-1 protein levels by ~60–70% (Extended Data Fig. 1). In contrast to a previous study⁹, the residual SYP-1 levels achieved using our experimental conditions were sufficient to permit assembly of SCs (albeit with reduced ratios of SYPs to axis subunits) and formation of chiasmata for all six chromosome pairs in most gonads (see Supplementary Discussion and Methods). Although SYPs are required to form crossovers, we found that partial depletion of SYP-1, SYP-2 or SYP-3 increases COSA-1 foci (Fig. 1a and Extended Data Fig. 2), indicating

that a sufficient pool of SYPs is required to limit COSA-1-marked crossovers to one per chromosome pair.

In principle, an increase in COSA-1-marked crossover sites could reflect an increase in the number of double-strand breaks (DSBs) formed and/or an increase in the fraction of DSBs repaired as crossovers. Therefore, we controlled DSB number by using γ -irradiation to induce DSBs in *spo-11* mutant worms, which are proficient for pairing and SC assembly but lack endogenous meiotic DSBs (Fig. 1b)¹⁴. Controls recapitulated the previous finding that increasing DSBs beyond the level needed to ensure at least one per chromosome pair (1,000 Rad, 23.5 DSBs per nucleus) did not increase the average number of COSA-1 foci per nucleus beyond six⁵. Further, COSA-1 foci exhibited very low standard deviations (± 0.14 – 0.27), reflecting operation of the robust crossover control system. By contrast, upon *syp-1* RNAi, most nuclei

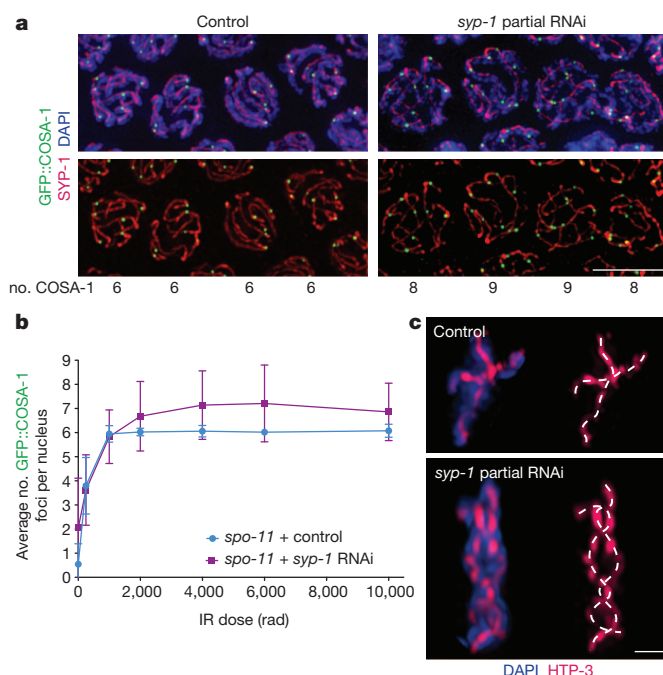


Figure 1 | SYP-1 partial depletion increases numbers of COSA-1 foci and chiasmata. **a**, Immunofluorescence images of late pachytene nuclei from control or *syp-1* partial RNAi worms. DAPI, 4',6-diamidino-2-phenylindole. Scale bar, 5 μ m. **b**, Dose–response graph depicting mean numbers of green fluorescent protein (GFP)::COSA-1 foci formed per nucleus in response to DSBs generated by increasing doses of γ -irradiation (IR, rad). See Methods for numbers of nuclei used; error bars indicate \pm s.d. At $>1,000$ rad, both the mean numbers of foci and s.d. were increased in *syp-1* RNAi relative to control. **c**, Three-dimensionally rendered images of individual diakinesis bivalents comprising the *mnt12* (X;IV) fusion chromosome pair. Dashed lines (white) indicate traced HTP-3 axes, with crossing of axes indicating chiasmata. Scale bar, 1 μ m.

¹Department of Developmental Biology, Stanford University, School of Medicine, Stanford, California 94305, USA. ²Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA. ³Department of Genetics, Stanford University, School of Medicine, Stanford, California 94305, USA.

exposed to $>1,000$ rad of γ -irradiation had >6 COSA-1 foci and standard deviations were much higher (± 1.2 – 1.6), indicating impairment of crossover control. This experiment shows that increased DSBs alone cannot account for the increase in COSA-1 foci following SYP-1 partial depletion, and that for any given level of DSBs, SYP-1 has a role in determining the fraction that will mature into cytologically differentiated crossovers.

Our data suggest that SYP-1 partial depletion impairs crossover interference. To investigate this further, we used worms homozygous for the two-chromosome fusion *mnT12* (X chromosome fused with chromosome IV)^{10,15}. Previous work showed that although crossover interference limits the *mnT12* chromosome pair to a single COSA-1 focus in the majority of meioses, *mnT12* chromosome pairs with two COSA-1 foci also occur⁵. Under our control conditions, 51% of *mnT12* pairs had a single COSA-1 focus, whereas 49% had two foci (Fig. 2a, b and Methods). This occurrence of two COSA-1 foci along *mnT12* allowed us to assess interference strength in the context of wild-type SYP-1 levels and compare it to crossover/chiasma interference in the context of SYP-1 partial depletion (Figs 1c and 2–4).

Increased COSA-1 foci following *syp-1* RNAi correlated with an increased number of cytologically resolvable chiasmata on *mnT12* bivalents at diakinesis, the last stage of prophase (Fig. 1c and Extended Data Fig. 3). Consistent with *mnT12* having only one or two COSA-1 foci at the late pachytene stage in controls (average of 1.49 COSA-1 foci per *mnT12* pair), control *mnT12* diakinesis bivalents had only one or two chiasmata. Upon *syp-1* RNAi, an average of 2.57 COSA-1 foci per *mnT12* pair were observed at late pachytene (49% having ≥ 3 foci; Fig. 2b), and 47% of diakinesis bivalents had ≥ 3 chiasmata. These and other data (Extended Data Fig. 3) indicate that the extra COSA-1 foci in *syp-1* RNAi worms represent bona fide cytologically differentiated inter-homologue crossovers.

We conducted quantitative analyses evaluating the positions, distributions and distances between COSA-1-marked sites on computationally straightened *mnT12* chromosomes (with HTP-3 immunofluorescence marking chromosome axes¹⁶ and with immunofluorescence for the HIM-8 protein, which localizes near the left end of the X chromosome¹⁷, serving as an orientation marker; Fig. 2a and Methods). The average distance between COSA-1 foci on *mnT12* with ≥ 2 COSA-1 foci in *syp-1* RNAi was $4.5 \mu\text{m}$, substantially shorter than the average distance of $8.5 \mu\text{m}$ for control *mnT12* with two foci ($P < 0.0001$). Further, even when only *syp-1* RNAi *mnT12* with two foci were considered, the average distance ($6.5 \mu\text{m}$) was still significantly shorter than in controls ($P < 0.0001$; Fig. 2c). These data show that *syp-1* RNAi decreases the effective distance over which interference operates. The average distance between COSA-1 foci on control *mnT12* ($8.5 \mu\text{m}$) exceeds the average axis length of unfused autosomes ($6.4 \mu\text{m}$, $P < 0.0001$), reinforcing the indirect inference from ref. 10 that interference in *C. elegans* operates over distances longer than the length of a normal chromosome.

We assessed crossover interference further by dividing the *mnT12* axis into eight evenly spaced intervals and binning the positions of COSA-1 foci into these intervals (Fig. 2d, e and Extended Data Fig. 4). First, we compared the distributions of COSA-1 foci among these intervals for chromosomes with different numbers of foci. In controls, COSA-1 foci exhibited a specific non-random distribution indicative of interference: for the subset of chromosomes with only one COSA-1 focus, the single focus usually occurred near the middle of *mnT12*, whereas for the subset of chromosomes with two COSA-1 foci, those foci predominantly occurred near the chromosome ends (Fig. 2d). By contrast, *syp-1* RNAi resulted in a fairly even distribution of foci along *mnT12*, regardless of the number of COSA-1 foci, consistent with attenuation of crossover interference. Second, plotting the separation between adjacent pairs of COSA-1 foci (Fig. 2e) revealed wide spacing (≥ 4 interval boundaries) separating adjacent COSA-1 foci in controls, and reduced separation between adjacent foci in *syp-1* RNAi worms, both when all pairs of foci or only chromosomes with two foci were considered.

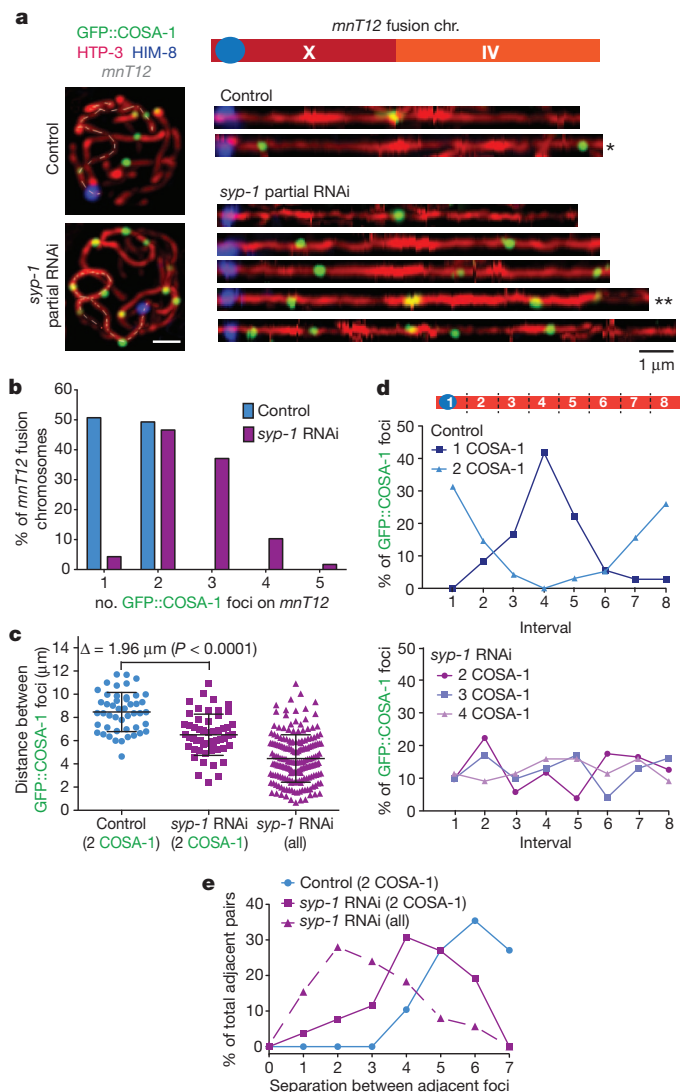


Figure 2 | SYP-1 partial depletion attenuates crossover interference. **a**, Left, projections of individual late pachytene nuclei containing the *mnT12* fusion chromosome (depicted schematically at top right), which is identified by its associated HIM-8 focus; grey dotted line indicates the traced three-dimensional path of the *mnT12* chromosome axis. Right, representative computationally straightened *mnT12* used for quantitative analyses in **b–e** and Figs 3 and 4. Asterisks indicate the straightened chromosomes from the nuclei shown. Scale bars, 1 μm . **b**, Graph indicating the percentage of *mnT12* with the indicated number of COSA-1 foci for control ($n = 69$) and *syp-1* RNAi ($n = 115$). **c**, Scatterplot showing measured distances (μm) between COSA-1 foci on control *mnT12* with two foci, on *syp-1* RNAi *mnT12* with two foci, and on all *syp-1* RNAi *mnT12* with multiple foci. Horizontal lines indicate the mean; error bars indicate s.d. **d**, Graphs indicating the distributions of foci among eight evenly spaced intervals along the *mnT12* axis (schematic at top) for the subsets of control (top graph) and *syp-1* RNAi (bottom graph) *mnT12* with the indicated numbers of COSA-1 foci. **e**, Graph depicting the distributions of spacing between adjacent pairs of foci. Using the same eight-interval scheme as in **d**, separation values are defined as the number of interval boundaries crossed before encountering the next focus (that is, adjacent foci within the same interval on the same chromosome pair have a separation of 0, adjacent foci in consecutive intervals have a separation of 1, etc.). See Methods for numbers of *mnT12* chromosomes used for **c–e**.

We also used two quantitative methods to calculate interference strength. First, we performed a coefficient of coincidence analysis, with *mnT12* divided into four intervals (Fig. 3a, Extended Data Fig. 4, Extended Data Table 1 and Methods). Controls exhibit a robust signature of interference (I): complete interference ($I = 1$) for adjacent interval pairs; strong but reduced interference for pairs at intermediate

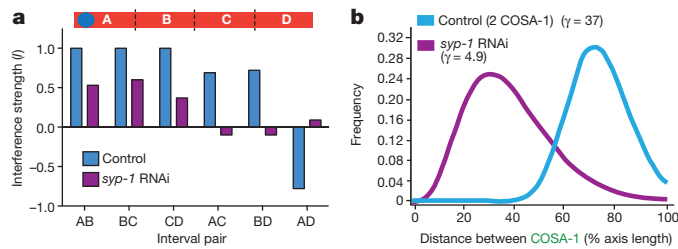


Figure 3 | SYP-1 partial depletion decreases crossover interference strength. **a**, Graph of interference strength (I) values for the indicated interval pairs, where $I = 1 - \text{observed/expected}$ (Methods); schematic (top) indicates division of *mnT12* into four intervals for this analysis. **b**, Graph showing best-fit probability density function curves generated when a gamma distribution was used to model the distribution of inter-focus distances; for this analysis, distances between adjacent foci were expressed as percentage of axis length. See Methods for numbers of chromosomes used.

distance; and high 'negative' interference ($I < 0$) for the pair of intervals including opposite ends of the chromosome, indicating that a chromosome with a focus in one end-interval has an increased likelihood of a second focus in the opposite end-interval (Fig. 2d, top). By contrast, the *sy-1* RNAi data show reduced interference for adjacent interval pairs, and no interference for interval pairs separated by ≥ 1 interval, indicating attenuated interference. Second, we used a gamma distribution to model inter-focus distances (expressed as a percentage of total axis length), and generated best-fit probability density curves where the shape parameter (γ) is a relative indicator of interference strength, with $\gamma = 1$ indicating no interference and higher values signifying stronger interference¹⁸ (Fig. 3b and Extended Data Fig. 5). Whereas the control displayed very strong interference ($\gamma = 37$), *sy-1* RNAi displayed a substantial attenuation of interference ($\gamma = 4.9$).

A separate analysis of our *mnT12* data yielded another key finding: COSA-1-marked crossovers are associated with a local increase in chromosome axis length (Fig. 4). Plotting the mean axis length for *sy-1* RNAi *mnT12* with 1, 2, 3 or 4 foci revealed a striking linear relationship between number of COSA-1 foci and mean axis length ($R^2 = 0.952$), with the slope of the linear regression line indicating that each COSA-1 focus is associated with a 0.4- μm increase in axis length (Fig. 4a). Further, association between number of foci and axis length was also observed in controls (Fig. 4a). Moreover, comparison of *mnT12* axis lengths in a *spo-11* mutant (which lacks crossovers) and in *spo-11/+* heterozygous controls yielded a similar linear relationship for *mnT12* with 0, 1 or 2 COSA-1 foci (Fig. 4b; slope = 0.49, $R^2 = 0.952$). Analysis of unfused autosomes from *spo-11*, controls and *sy-1* RNAi worms likewise showed each COSA-1 focus associating with a 0.4- μm increase in axis length (Fig. 4c; $R^2 = 0.999$), demonstrating that the relationship between crossovers and axis length is generalizable to other chromosomes. Together, these data indicate that extension of chromosome axes occurs in response to crossover designation. Finally, assessment of the distance between the left end of the chromosome and the position of the HIM-8 focus on control *mnT12* (Fig. 4d and Extended Data Fig. 6) revealed that the mean length of this specific short axis segment was increased by 0.4–0.5 μm when a COSA-1 focus was present in the segment. As this value is comparable to that inferred from the linear regression analyses, we conclude that the increase in axis length associated with each crossover is predominantly a local effect.

Our demonstration that partial depletion of SYPs attenuates the robust crossover interference in *C. elegans* indicates a role for SC central region proteins in crossover control. This effect of SYP depletion on interference could reflect a decreased ability to propagate an inhibitory signal, a reduced sensitivity of recombination precursors to inhibition, and/or a prolonged state of competence for crossover designation. Our finding that SC central region subunits have a role in achieving the high level of interference characteristic of *C. elegans* meiosis can be reconciled with studies concluding that the SC is dispensable for

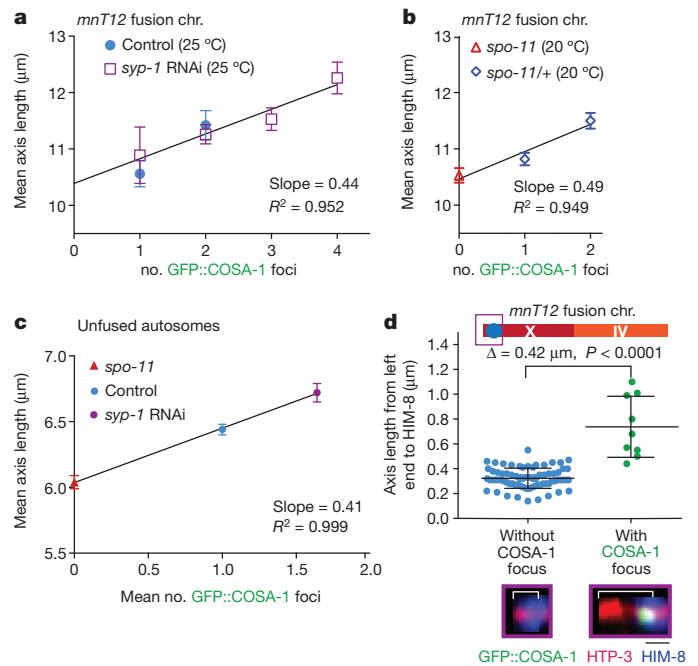


Figure 4 | Crossover designation causes a local expansion of chromosome axis length. **a**, **b**, Graphs plotting the relationship between number of COSA-1 foci and mean length (\pm s.e.m.) of the *mnT12* chromosome axis. **a**, Data for *sy-1* RNAi and control at 25 °C; extrapolated linear regression line was generated using the *sy-1* RNAi data. **b**, Data for *spo-11* and *spo-11/+* control at 20 °C. **c**, Graph plotting the relationship between mean number of COSA-1 foci and mean axis length (\pm s.e.m.) for unfused autosomes in *spo-11*, control and *sy-1* RNAi nuclei (Methods). **d**, Scatterplot showing length measurements (from control nuclei at 25 °C) for the axis segment extending from left end of *mnT12* to the centre of the HIM-8 focus for chromosomes that either lacked or had a COSA-1 focus in this segment, as depicted in images below the graph. Horizontal lines indicate mean; error bars indicate s.d. See Methods for numbers of chromosomes used.

interference in budding yeast^{19–21}, on the basis of several considerations. There is a growing body of evidence that, at least in some organisms, non-random distribution of prospective crossover intermediates may occur in (at least) two different steps^{22–24}. This suggests that multiple layers of regulation can contribute to the final interference distribution of crossovers, raising the possibility that the relative contributions of different crossover control mechanisms may differ between organisms. Thus, given that the attenuated interference observed following SYP-1 partial depletion in *C. elegans* ($\gamma = 4.9$) appears at least as strong or stronger than the interference observed in wild-type budding yeast ($\gamma = 1.9$)²⁵, we suggest that *C. elegans* may use a layer of crossover control that does not make an appreciable contribution in budding yeast.

Integrating our findings with previous data showing that crossover designation in *C. elegans* occurs in the context of assembled SCs^{5,26}, we suggest a model in which crossover designation requires a capacity for local expansion of previously assembled meiotic chromosome structures. We suggest that local expansion in turn alters the thickness, density and/or rigidity of structures beyond the local area, thereby reducing the capacity for axis expansion at other sites (see Supplementary Discussion). Overall, our work supports the idea that meiotic crossover regulation operates as a self-limiting system in which meiotic chromosome structures create an environment that promotes crossovers, which in turn modify chromosome structures to inhibit crossover formation at additional neighbouring sites.

METHODS SUMMARY

For assessing the effects of partial depletion of SYP-1, worms were exposed to RNAi treatment at 25 °C from the L1 stage until 16–22 h post-L4 stage, when their gonads were then dissected and processed for immunofluorescence. For Fig. 1,

chromosomes in three-dimensionally preserved meiotic prophase nuclei were imaged using wide-field deconvolution microscopy. For analyses involving measurements of chromosome axis lengths and positions of GFP::COSA-1 foci (Figs 2–4), three-dimensionally preserved nuclei were imaged using confocal microscopy, followed by three-dimensional tracing and subsequent computational straightening of chromosomes before recording of position coordinates²⁷. All *P* values reported in the main text are two-tailed and calculated from Mann–Whitney tests.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 April; accepted 16 August 2013.

Published online 9 October 2013.

- Page, S. L. & Hawley, R. S. Chromosome choreography: the meiotic ballet. *Science* **301**, 785–789 (2003).
- Martinez-Perez, E. & Colaiacovo, M. P. Distribution of meiotic recombination events: talking to your neighbors. *Curr. Opin. Genet. Dev.* **19**, 105–112 (2009).
- Muller, H. J. The mechanism of crossing-over. *Am. Nat.* **50**, 193–221 (1916).
- Sturtevant, A. H. The linear arrangements of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* **14**, 43–59 (1913).
- Yokoo, R. *et al.* COSA-1 reveals robust homeostasis and separable licensing and reinforcement steps governing meiotic crossovers. *Cell* **149**, 75–87 (2012).
- Bishop, D. K. & Zickler, D. Early decision: meiotic crossover interference prior to stable strand exchange and synapsis. *Cell* **117**, 9–15 (2004).
- Egel, R. Synaptonemal complex and crossing-over: structural support or interference? *Heredity* **41**, 233–237 (1978).
- Maguire, M. P. Can abortive early homologous associations promote increased crossing-over in an adjacent rearranged segment? *Genome* **30**, 469–472 (1988).
- Hayashi, M., Mlynarczyk-Evans, S. & Villeneuve, A. M. The synaptonemal complex shapes the crossover landscape through cooperative assembly, crossover promotion and crossover inhibition during *Caenorhabditis elegans* meiosis. *Genetics* **186**, 45–58 (2010).
- Hillers, K. J. & Villeneuve, A. M. Chromosome-wide control of meiotic crossing over in *C. elegans*. *Curr. Biol.* **13**, 1641–1647 (2003).
- Colaiacovo, M. P. *et al.* Synaptonemal complex assembly in *C. elegans* is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. *Dev. Cell* **5**, 463–474 (2003).
- MacQueen, A. J., Colaiacovo, M. P., McDonald, K. & Villeneuve, A. M. Synapsis-dependent and -independent mechanisms stabilize homolog pairing during meiotic prophase in *C. elegans*. *Genes Dev.* **16**, 2428–2442 (2002).
- Smolnikov, S. *et al.* SYP-3 restricts synaptonemal complex assembly to bridge paired chromosome axes during meiosis in *Caenorhabditis elegans*. *Genetics* **176**, 2015–2025 (2007).
- Dernburg, A. F. *et al.* Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell* **94**, 387–398 (1998).
- Sigurdson, D. C., Herman, R. K., Horton, C. A., Kari, C. K. & Pratt, S. E. An X-autosome fusion chromosome of *Caenorhabditis elegans*. *Mol. Gen. Genet.* **202**, 212–218 (1986).
- Goodyer, W. *et al.* HTP-3 links DSB formation with homolog pairing and crossing over during *C. elegans* meiosis. *Dev. Cell* **14**, 263–274 (2008).
- Phillips, C. M. *et al.* HIM-8 binds to the X chromosome pairing center and mediates chromosome-specific meiotic synapsis. *Cell* **123**, 1051–1063 (2005).
- McPeck, M. S. & Speed, T. P. Modeling interference in genetic recombination. *Genetics* **139**, 1031–1044 (1995).
- Fung, J. C., Rockmill, B., Odell, M. & Roeder, G. S. Imposition of crossover interference through the nonrandom distribution of synapsis initiation complexes. *Cell* **116**, 795–802 (2004).
- Börner, G. V., Kleckner, N. & Hunter, N. Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* **117**, 29–45 (2004).
- Storlazzi, A., Xu, L., Schwacha, A. & Kleckner, N. Synaptonemal complex (SC) component Zip1 plays a role in meiotic recombination independent of SC polymerization along the chromosomes. *Proc. Natl Acad. Sci. USA* **93**, 9043–9048 (1996).
- Cole, F. *et al.* Homeostatic control of recombination is implemented progressively in mouse meiosis. *Nature Cell Biol.* **14**, 424–430 (2012).
- de Boer, E., Stam, P., Dietrich, A. J., Pastink, A. & Heyting, C. Two levels of interference in mouse meiotic recombination. *Proc. Natl Acad. Sci. USA* **103**, 9607–9612 (2006).
- Reynolds, A. *et al.* RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. *Nature Genet.* **45**, 269–278 (2013).
- Chen, S. Y. *et al.* Global analysis of the meiotic crossover landscape. *Dev. Cell* **15**, 401–415 (2008).
- Rosu, S., Libuda, D. E. & Villeneuve, A. M. Robust crossover assurance and regulated interhomolog access maintain meiotic crossover number. *Science* **334**, 1286–1289 (2011).
- Mets, D. G. & Meyer, B. J. Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. *Cell* **139**, 73–86 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Dernburg and M. Zetka for antibodies and the CGC (funded by National Institutes of Health (NIH) P40 OD010440) for strains. We thank K. Hillers and K. Zawadzki for comments on the manuscript. This work was supported by a Helen Hay Whitney Foundation Postdoctoral Fellowship, a Katharine McCormick Advanced Postdoctoral Fellowship, and NIH K99 HD076165 to D.E.L. and by NIH R01 GM067268 to A.M.V. B.J.M. is an investigator of the Howard Hughes Medical Institute.

Author Contributions D.E.L. and A.M.V. conceived and designed the experiments, analysed the data and wrote the paper. D.E.L. performed the experiments. S.U. deconvolved confocal images. S.U. and B.J.M. provided scientific discussions and technical expertise for computational straightening of chromosomes.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.V. (annev@stanford.edu).

METHODS

C. elegans strains, genetics and culture conditions. All strains are from the Bristol N2 background and were maintained and crossed at 20 °C under standard conditions. Temperatures used for specific experiments are indicated in Figures and below.

The following strains were used in this study: (1) AV307: *sy-1(me17) V/nT1 [unc-?(n754) let-? qIs50] (IV;V)*. (2) AV630: *meIs8[unc-119(+)] Ppie-1::gfp::cosa-1] II*. (3) AV647: *meIs8[unc-119(+)] Ppie-1::gfp::cosa-1] II; spo-11(me44) IV/nT1 [unc-?(n754) let-? qIs50] (IV;V)*. (4) AV695: *meIs8[unc-119(+)] Ppie-1::gfp::cosa-1] II; mnT12 (IV;X)*. (5) AV760: *meIs8[unc-119(+)] Ppie-1::gfp::cosa-1] II; unc-24(e138) dpy-4(e1166) mnT12 (IV;X)*. (6) AV761: *meIs8[unc-119(+)] Ppie-1::gfp::cosa-1] II; spo-11(me44) IV/nT1[qIs51] (IV;V)*. Additional information on strains: *qIs50* contains [*Pmyo-2::gfp; Ppes-10::gfp; PF22B7.9::gfp*] and *qIs51* contains [*Pmyo-2::gfp; Ppes-10::gfp; PF22B7.9::gfp*].

For Fig. 4b, c and Extended Data Fig. 6, *meIs8; spo-11 mnT12* worms (and control *meIs8; spo-11/+ mnT12* worms) were obtained by the following scheme: (1) AV761 males × AV760 hermaphrodites; (2) *meIs8; + spo-11 + IV / unc-24(e138) + dpy-4(e1166) mnT12 (IV;X)* males (from cross 1) × AV760 hermaphrodites; (3) progeny from cross 2 were singly plated. *meIs8; + spo-11 + mnT12 / unc-24(e138) + dpy-4(e1166) mnT12 (IV;X)* hermaphrodites were identified on the basis of failure to produce male self progeny, indicating homozygosity for *mnT12*; (4) from step 3 plates that did not produce male progeny, 20–24 h post-L4 non-Unc non-Dpy hermaphrodite progeny were used for immunofluorescence. These included both *spo-11* homozygotes and *spo-11/+* worms, which were used as controls. *spo-11* mutant and control gonads were dissected onto the same slides and were differentiated during imaging based on DAPI-stained bodies in diakinesis-stage oocytes (10 univalents in *spo-11* oocytes and 5 bivalents in controls). **Experimental conditions for partial depletion of SYP-1 by RNAi.** For *sy-1* partial RNAi, synchronized L1 larvae (grown at 20 °C from eggs hatched on unseeded plates for 24 h) were placed on fresh NGM + Amp + isopropyl β-D-1-thiogalactopyranoside (IPTG) plates recently seeded with *Escherichia coli* HT115 cells containing either a fragment of the *sy-1/F26D2.2* gene in the L4440 vector, or the empty vector (referred to as control) from the Ahringer laboratory RNAi library²⁸. For analyses of GFP::COSA-1 foci in late pachytene nuclei, hermaphrodite worms were grown at 25 °C starting at the L1 stage and dissected for immunofluorescence at 16–22 h post-L4 stage to obtain optimal SYP-1 depletion for observing an increase in GFP::COSA-1 foci; for quantification of chiasmata, worms were dissected at 24–30 h post-L4 to permit nuclei with optimal SYP-1 depletion in late pachytene at 16–22 h post-L4 to progress to diakinesis. Differences in experimental culture conditions (that is, growth at 25 °C on *E. coli* HT115 on NGM + Amp + IPTG versus growth at 20 °C on *E. coli* OP50 on NGM) likely account for the higher baseline frequency of *mnT12* chromosome pairs with two GFP::COSA-1 foci in the present study (49%) compared to a previous study (30%)⁵.

Evaluation of SYP-1 partial depletion. RNAi partial depletion conditions used for our experiments were consistently and deliberately much less severe than the level of partial SYP-1 depletion used in ref. 9. Under the conditions used for the current work, the vast majority of pachytene nuclei had assembled SCs on all six chromosome pairs and most diakinesis oocytes had chiasmata connecting all six chromosome pairs (see main text). (By contrast, the vast majority of pachytene nuclei analysed in ref. 9 had assembled SCs on only 1, 2 or 3 chromosome pairs, and in the worms used for their crossover analysis, most diakinesis oocytes had multiple achiasmate chromosome pairs⁹.) To assess the extent of partial depletion used in the current study, two different approaches were used, immunoblotting and fluorescence quantification:

Immunoblotting: the extent of SYP-1 depletion achieved by our RNAi conditions was estimated using western blot analysis (Extended Data Fig. 1a). Protein lysates were generated from control, *sy-1* RNAi and *sy-1(me17)* mutant worms (100 worms each; see below). For each experiment, protein blots containing identical amounts of each lysate, plus a dilution series of the control lysate, were probed with α-SYP-1 antibody, and the extent of depletion was estimated by comparing the residual SYP-1 band in the *sy-1* RNAi lysate to the SYP-1 bands in the dilution series from a control lysate processed in parallel.

Protocol details: 100 worms (treated and staged as for the reported immunofluorescence analyses) were picked to unseeded plates, washed off plates with 1.5 ml of cold M9, and spun down at 2,000 r.p.m. for 1 min. After placing on ice for 1 min, supernatant was removed, and worm pellets were washed three times with cold M9 before the addition of 2 × Laemmli buffer with β-mercaptoethanol (Bio-Rad) and boiling for 10 min to create worm protein lysates. Samples of each lysate were run on a 4–15% Bio-Rad TGX acrylamide gel and then transferred to nitrocellulose membrane using a wet Western blot transfer system (Bio-Rad). Membrane was blocked in 5% milk in 1 × Tris Buffered Saline with Tween 20 (TBST) for 1 h at room temperature (approx. 22–23 °C) and then incubated with 1:500 dilution of

rabbit anti-SYP-1 antibody¹² (pre-absorbed with *sy-1(me17)* dissected worms) in 5% milk in 1 × TBST overnight at 4 °C. Membrane was washed 5 × 10 min in 1 × TBST before addition of 1:4,000 dilution of goat anti-rabbit HRP-conjugated antibody in 5% milk in TBST for 2 h at room temperature. Following 5 × 10-min washes in 1 × TBST, membrane was incubated with Clarity Western ECL substrate (Bio-Rad) for 5 min at room temperature and then exposed to film. To assess loading, antibodies were removed from the membrane with mild stripping buffer (Abcam), and the membrane was then blocked, incubated and washed as described above except for the following modification for the primary antibody: 1:10,000 mouse anti-α-tubulin antibody (Sigma no. T6199) in 5% milk in TBST, incubated for 1 h at 4 °C.

Fluorescence quantification: for Extended Data Fig. 1b, c, normalized fluorescence levels were assessed from immunofluorescence slides of control and *sy-1* partial RNAi gonads imaged on the DeltaVision microscope using the imaging conditions described below. During image acquisition, exposure times for channels assessing SYP-1 and HTP-3 fluorescence signals were kept the same for both control and *sy-1* RNAi slides. Deconvolved images were projected using a sum projection algorithm with softWoRx (Applied Precision) software. Using the Edit Polygon tool in softWoRx, individual nuclei from a specific region of the gonad (from two rows before COSA-1 foci formation, to three rows following COSA-1 foci formation) were traced and assessed for SYP-1 and HTP-3 fluorescence intensity. Nuclei were assessed only if they were completely contained within the image stack and did not overlap with any other nuclei in the projected image. (Nuclei with compact DAPI signals indicative of arrest and/or apoptosis were excluded.) Further, a subset of *sy-1* RNAi gonads (40%) that exhibited severe SYP-1 depletion (that is, SYP-1 was not detected on all chromosomes) was excluded from this analysis, as gonads in this category would have been excluded from our experiments evaluating numbers and distribution COSA-1 foci on the basis of presence of asynapsed chromosomes and/or inability to trace continuous chromosome axes (as a result of asynapsis).

For each gonad, background fluorescence in the SYP-1 and HTP-3 channels was assessed by measuring fluorescence intensities for three separate areas between nuclei in the scored region and using these to calculate the average background fluorescence per pixel. For each individual nucleus, SYP-1 and HTP-3 fluorescence measurements were calculated by dividing the total intensity by the total area of the projected nucleus, and then subtracting the average background fluorescence for the gonad. To obtain an average SYP-1:HTP-3 ratio for each gonad, the SYP-1:HTP-3 ratios calculated for each nucleus within a given gonad were averaged. Once average SYP-1:HTP-3 ratios were measured for all gonads, two different approaches were used to compare SYP-1:HTP-3 ratios for *sy-1* RNAi gonads and control gonads: (1) ‘mean of experiments’: for each experiment, the average SYP-1:HTP-3 ratios for control gonads and for *sy-1* RNAi gonads were calculated; the control value was set to 1, and the normalized SYP-1:HTP-3 ratio for *sy-1* RNAi for that particular experiment was expressed as a fraction of the control value. The corresponding graph plots the average normalized SYP-1:HTP-3 ratio for *sy-1* RNAi ± s.d. for three independent experiments; (2) ‘mean of all gonads’: for each individual experiment, the SYP-1:HTP-3 ratio for each individual gonad (both *sy-1* RNAi and control) was normalized to the average control SYP-1:HTP-3 ratio obtained from the same experiment. These values were then used to calculate average (± s.d.) normalized SYP-1:HTP-3 ratios for all control gonads and for all *sy-1* RNAi gonads across all experiments. Note that this latter approach makes it possible to convey the variability in the control measurements in the corresponding graph.

Numbers of gonads assessed: experiment no. 1: control = 7 gonads, *sy-1* RNAi = 7 gonads; experiment no. 2: control = 15 gonads, *sy-1* RNAi = 6 gonads; experiment no. 3: control = 10 gonads, *sy-1* RNAi = 8 gonads.

Of the two approaches, the western blot analysis provides a somewhat lower estimate of residual SYP-1 levels, in part because the subset of worms with more severe SYP-1 depletion (which would have been excluded from our experiments) are included in the protein lysates. Thus, the immunofluorescence approach probably provides a better estimate of the residual SYP-1 levels present in the nuclei analysed in our experiments.

Because our data indicate that a 60–70% reduction in SYP-1 levels can lead to increased COSA-1-marked crossover sites and impaired interference, we also tested whether the *sy-1* locus might be haploinsufficient. As we observed 6.02 ± 0.31 (mean ± s.d., $n = 53$) and 6.00 ± 0.13 (± s.d., $n = 114$) GFP::COSA-1 foci per late pachytene nucleus of *sy-1(me17)/+* and *sy-3(ok758)/+* worms, respectively, we conclude that reducing *sy-1* gene dose by half is not sufficient to impair the robust crossover control system.

Immunofluorescence. Immunofluorescence experiments were performed as described previously²⁹, with modifications. Gonads from adult worms at 18–24 h post-L4 stage were dissected. Slides were mounted with Vectashield (except for confocal images, which were mounted with Invitrogen ProLong Gold) and a

coverslip with a $170 \pm 5\text{-}\mu\text{m}$ thickness. Slides were sealed with nail polish immediately following mounting and stored at 4°C before imaging (except for confocal slides which were cured for 24 h at room temperature before sealing and storing at 4°C). All slides were imaged (as described below) within 2 weeks of preparation. The following primary antibody dilutions were used: rabbit anti-GFP (1:1,000)²⁵; chicken anti-GFP (1:1,000) (Abcam 13970); guinea pig anti-SYP-1 (1:200)¹²; chicken anti-HTP-3 (1:500)³⁰; rabbit anti-HIM-3 (1:200)³¹; guinea pig anti-HIM-8 (1:500)¹⁷.

Imaging. For Fig. 1, immunofluorescence slides were imaged at 512×512 pixel dimensions on an Applied Precision DeltaVision microscope using a $60\times$ objective with $1.5\times$ optivar. Images were acquired as Z-stacks at $0.2\text{-}\mu\text{m}$ intervals and deconvolved with Applied Precision softWoRx deconvolution software. For Figs 2–4, immunofluorescence slides were imaged at $1,024 \times 1,024$ pixel dimensions on a Leica SP2 AOBs Confocal microscope using a $63\times$ objective with $4.7\times$ zoom in 12-bit format. Images were acquired as Z-stacks at 81-nm intervals and deconvolved with Huygens Professional deconvolution software (Scientific Volume Imaging). For quantification of GFP::COSA-1 foci, nuclei that were in the last 4–5 rows of late pachytene and were completely contained within the image stack were analysed. Foci were quantified manually from deconvolved three-dimensional stacks. For visualization and quantification of chiasmata (Fig. 1c and Extended Data Fig. 3a), individual *mnT12* bivalents from diakinesis nuclei in -2 , -3 or -4 oocytes were identified on the basis of size, cropped, and rotated in three dimensions using Velocity (PerkinElmer) three-dimensional rendering software. Scoring of chiasmata was primarily on the basis of HTP-3 (chromosome axis) and DAPI staining, as GFP::COSA-1 dissociates from chromosomes during progression through the diakinesis stage. For Fig. 1a, images shown are projections through three-dimensional data stacks encompassing whole nuclei, generated with a maximum-intensity algorithm with the softWoRx (Applied Precision) software. For Fig. 1c, the images shown are snapshots of a Velocity three-dimensional rendering of individual diakinesis bivalents with maximum intensity rendering for HTP-3. Images shown in Figs 2a and 4c are projections through three-dimensional data stacks encompassing whole nuclei, or straightened chromosomes, generated with a maximum-intensity algorithm using Prism/IVE software³².

γ -irradiation experiments. L1 progeny from AV647 worms (heterozygous for *spo-11(me44)* and the *nT1* balancer chromosome) were subjected to RNAi conditions described earlier. *spo-11(me44)* homozygotes were picked at the L4 stage from the RNAi plates on the basis of a lack of dominant markers associated with the balancer and placed on freshly seeded RNAi plates at 25°C . At 12 h post-L4, 25°C *syp-1* and control RNAi-treated worms were γ -irradiated with a Cs-137 source at either 250 rad, 1,000 rad, 2,000 rad, 4,000 rad, 6,000 rad or 10,000 rad. Following γ -irradiation, worms were placed back at 25°C for 8 h until dissection for immunofluorescence. Unirradiated controls were kept at 25°C and dissected at the same time as γ -irradiated worms.

Before quantification of COSA-1 foci, we used SYP-1 immunofluorescence to evaluate the efficacy of SYP-1 partial depletion. In the majority of *syp-1* RNAi gonads (57%), pachytene nuclei had detectable SYP-1 along all six chromosome pairs, but immunofluorescence signals were reduced in intensity relative to controls and there were some small chromosomal regions that lacked SYP-1; gonads in this category ($n = 84$) were used for quantification of COSA-1 foci. Gonads with SYP-1 staining comparable to wild type (18%) and gonads with more severe SYP-1 depletion (that is, SYP-1 was not detected on all chromosomes, 25%) were excluded from this analysis.

For the quantification of COSA-1, the numbers of nuclei counted for each γ -irradiation dose are as follows: control: 0 rad ($n = 189$), 250 rad ($n = 310$), 1,000 rad ($n = 243$), 2,000 rad ($n = 201$), 4,000 rad ($n = 119$), 6,000 rad ($n = 212$), 10,000 rad ($n = 52$); *syp-1* partial RNAi: 0 rad ($n = 138$), 250 rad ($n = 170$), 1,000 rad ($n = 278$), 2,000 rad ($n = 278$), 4,000 rad ($n = 195$), 6,000 rad ($n = 280$), 10,000 rad ($n = 345$).

Measurements along straightened chromosomes and interference analyses. For quantitative analyses of axis lengths and positions and distributions of COSA-1 foci, *mnT12* chromosomes (identified by HIM-8 binding) and unfused autosomes (from the same nuclei) with continuous and unambiguous chromosome axis staining by either HTP-3 or HIM-3 immunofluorescence were traced in three dimensions through the centre of the chromosome axis (indicated by strongest staining) and computationally straightened as described before²⁷. Using these straightened chromosomes, the lengths of the chromosome axes and the positions of the HIM-8 focus and COSA-1 foci along *mnT12* were measured using Prism/IVE. The left end of the *mnT12* fusion chromosome (identified based on proximity to the HIM-8 focus) was assigned the zero coordinate, and position coordinates (in μm) were determined and recorded for the centre of the HIM-8 focus, the centres of each COSA-1 focus, and the opposite end of the chromosome axis.

For these analyses, only nuclei in the last four rows of late pachytene that were completely contained within the image stacks were analysed. For the *syp-1* RNAi

gonads analysed, *mnT12* was examined in all nuclei meeting these criteria (115 total *mnT12* chromosomes from 16 gonads, 4 experiments, except for Figs 2d, e and 3, where six *mnT12* chromosomes were omitted from the analyses owing to lack of positional information for proper interval binning). For control analyses where the relative frequencies of *mnT12* chromosomes with 1 versus 2 COSA-1 foci were a relevant factor (Figs 2b, 3a, 4a and Extended Data Fig. 4), we similarly used data from gonads where all nuclei meeting the above-specified criteria were included (69 *mnT12* chromosomes from 6 gonads, 3 experiments, except for Fig. 3a, where one *mnT12* chromosome was omitted from the analyses owing to lack of positional information for proper interval binning). For Figs 2c–e, 3b and 4d, 16 additional control *mnT12* chromosomes with 2 COSA-1 foci from 2 additional gonads were also included in the analyses. Numbers for Fig. 4b: *spo-11/+* control (93 *mnT12* chromosomes from 13 gonads, 5 experiments); *spo-11* (75 *mnT12* chromosomes from 7 gonads, 3 experiments). For Extended Data Fig. 6, data from the same nuclei, plus 20 additional *mnT12* control chromosomes with HIM-8 positional data (from 5 gonads), were included in the analysis. The best-fit linear regression lines in Fig. 4a, b (and corresponding R^2 values) were generated by the Prism graphing program (GraphPad Software). For Fig. 2d, e and Extended Data Fig. 4, positions of foci were normalized by dividing each chromosome into eight intervals of equal size and binning the foci into these eight intervals.

For the interference strength analysis in Fig. 3a, positions of foci were normalized by dividing each chromosome into four intervals of equal size and binning the foci into these 4 intervals. For each specified pair of intervals tested, this coefficient of coincidence analysis compared the observed frequency of COSA-1 foci occurring in both intervals with the frequency expected if foci occurred independently in the two intervals (that is, absence of interference), with interference strength (I) calculated as $(1 - \text{observed/expected})$. The 'expected' number of chromosomes with COSA-1 foci occurring in both of a given pair of intervals (X and Y) was calculated as: (measured frequency of COSA-1 foci in interval X) \times (measured frequency of COSA-1 foci in interval Y) \times (number of *mnT12* chromosomes examined).

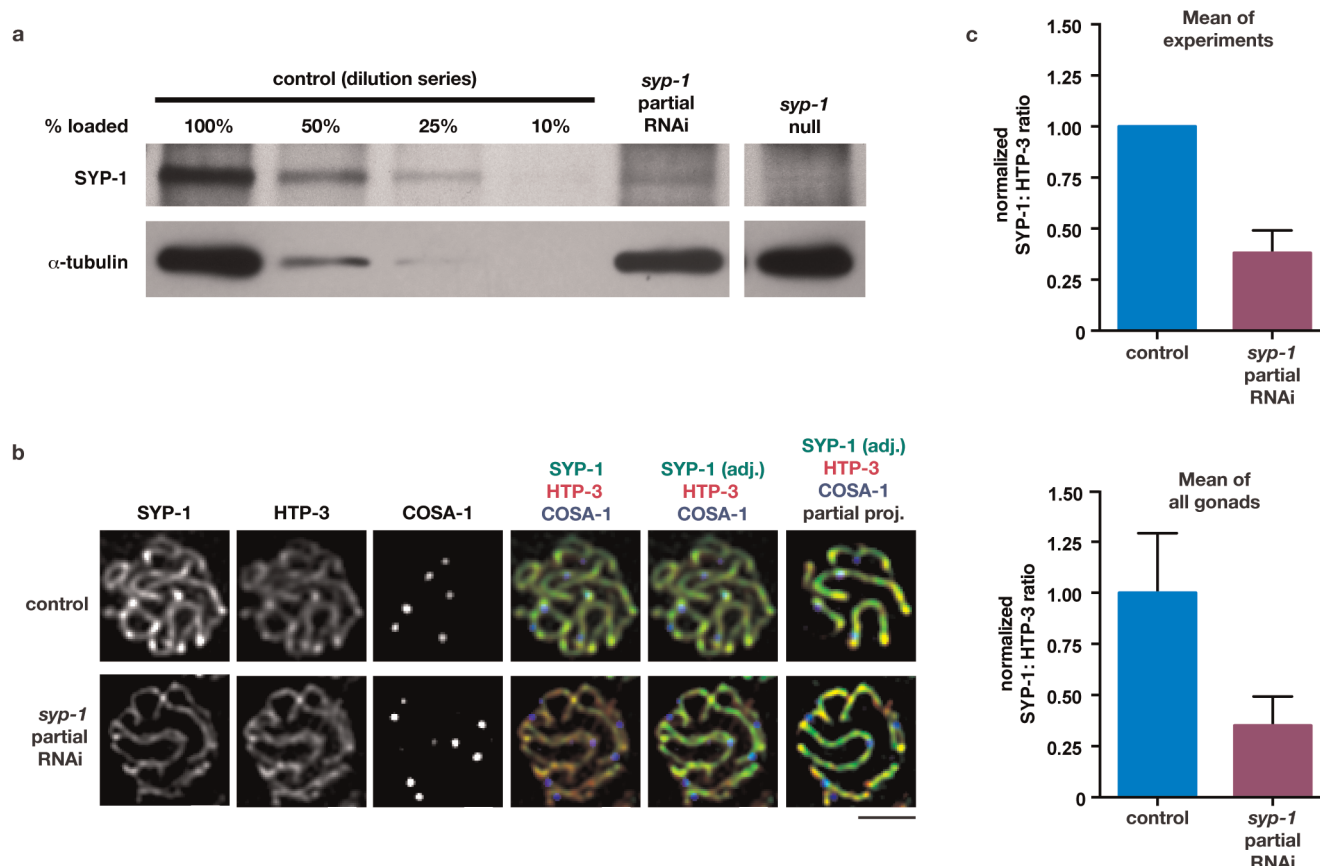
For determining interference strength using the best-fit gamma distribution analyses in Fig. 3b and Extended Data Fig. 5, distances between COSA-1 foci along straightened *mnT12* chromosomes were normalized by calculating the percent axis length separating neighbouring COSA-1 foci. Distances between adjacent foci expressed as per cent axis length were entered into the EasyFit Distribution Fitting Software (Mathwave) to generate histograms of the data sets and best-fit probability distribution curves. Numbers of inter-focus distances used are: control, $n = 47$; *syp-1* RNAi, $n = 183$.

In addition to measuring *mnT12* axis lengths, we also measured axis lengths of unfused autosomes with continuous and unambiguous HTP-3 or HIM-3 staining from the same nuclei used to measure *mnT12*. Numbers of chromosomes analysed were: controls at 25°C (208 autosomes from 62 nuclei, 6 gonads, 3 experiments); *syp-1* RNAi at 25°C (155 autosomes from 58 nuclei, 16 gonads, 4 experiments); *spo-11/+* controls at 20°C (233 autosomes from 77 nuclei, 13 gonads, 5 experiments); and *spo-11* at 20°C (203 autosomes from 62 nuclei, 7 gonads, 3 experiments). As mean axis lengths of unfused autosomes in RNAi controls at 25°C and *spo-11/+* controls at 20°C were nearly identical ($6.42 \pm 0.06\text{ }\mu\text{m}$ versus $6.46 \pm 0.04\text{ }\mu\text{m}$; \pm s.e.m.), these two data sets were combined to give a single control value for the graph in Fig. 4c.

Statistics. All P values reported in the main text and Extended Data Figs 2 and 6 are two-tailed and calculated from Mann–Whitney tests, which are robust non-parametric statistical tests appropriate for the relevant data sets. For the four-interval data set analysis in Fig. 3a, we used Fisher's exact test (an appropriate test for 2×2 contingency tables) to evaluate (for each pair of intervals) whether COSA-1 foci occurred independently in the two intervals under consideration (Extended Data Table 1). For the two-tailed P value reported in Extended Data Fig. 3b comparing the observed versus expected outcome for GFP::COSA-1 foci within an asynapsed region of the *mnT12* chromosome for *syp-1* RNAi, the χ^2 test was used as it is an appropriate test to compare observed versus expected outcomes for categorical data. For Fig. 4a–c, the goodness-of-fit for the linear regression lines were reported as coefficient of determination (R^2 ; an appropriate goodness-of-fit statistic for fitted linear regression lines), using the Prism graphing program (GraphPad Software). For all tests described above, the assumptions of each test were met.

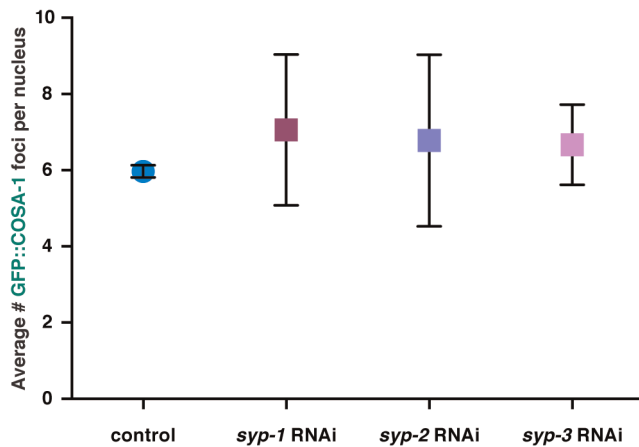
28. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
29. Martinez-Perez, E. & Villeneuve, A. M. HTP-1-dependent constraints coordinate homolog pairing and synapsis and promote chiasma formation during *C. elegans* meiosis. *Genes Dev.* **19**, 2727–2743 (2005).
30. MacQueen, A. J. *et al.* Chromosome sites play dual roles to establish homologous synapsis during meiosis in *C. elegans*. *Cell* **123**, 1037–1050 (2005).

31. Zetka, M. C., Kawasaki, I., Strome, S. & Muller, F. Synapsis and chiasma formation in *Caenorhabditis elegans* require HIM-3, a meiotic chromosome core component that functions in chromosome segregation. *Genes Dev.* **13**, 2258–2270 (1999).
32. Chen, H., Hughes, D. D., Chan, T. A., Sedat, J. W. & Agard, D. A. IVE (Image Visualization Environment): a software platform for all three-dimensional microscopy applications. *J. Struct. Biol.* **116**, 56–60 (1996).

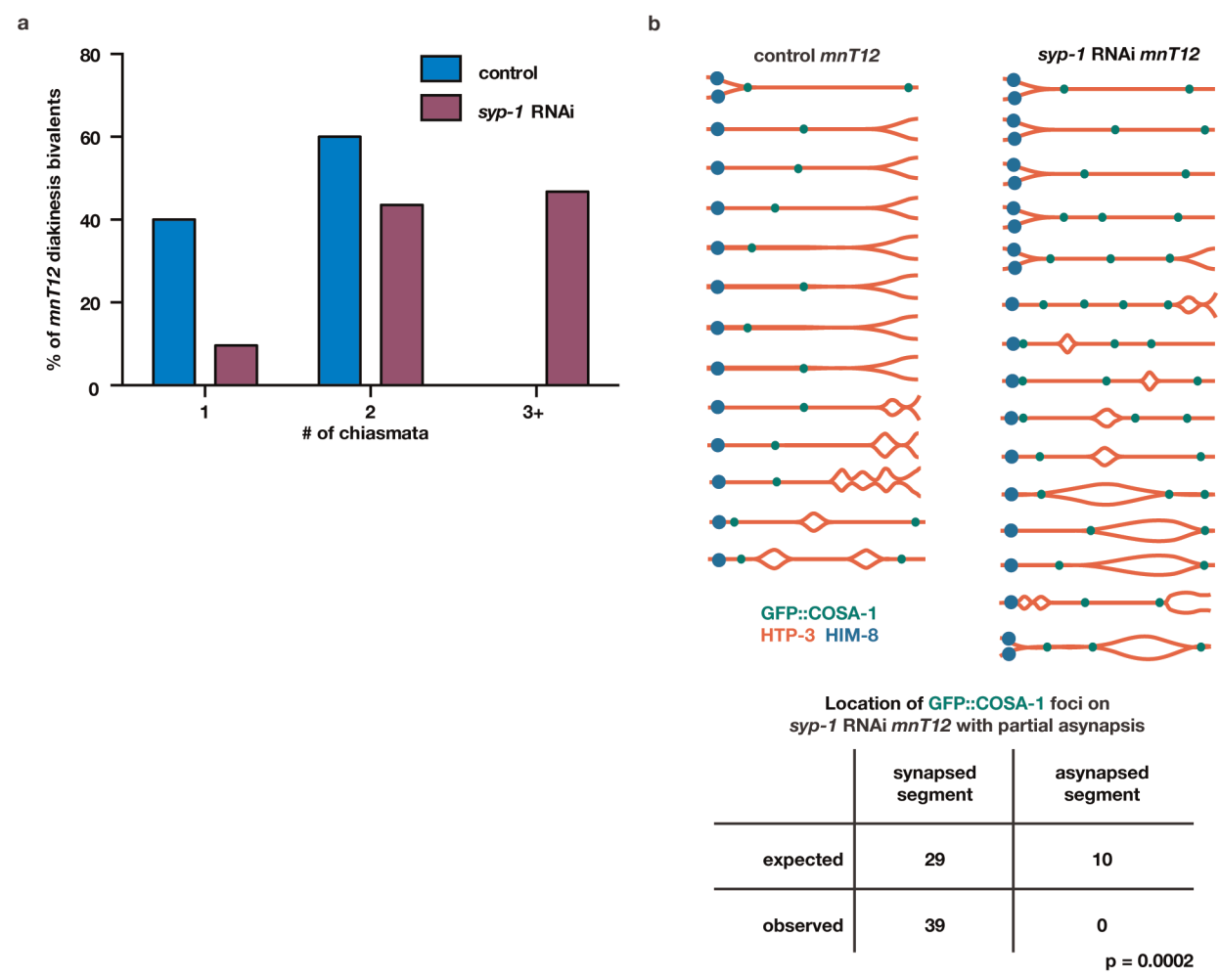


Extended Data Figure 1 | Quantification of SYP-1 partial depletion by attenuated RNAi. **a**, Representative western blot analysis of protein lysates from control, *syp-1* partial RNAi and *syp-1* null worms. A dilution series of control samples was used to estimate that the level of SYP-1 protein was reduced to approximately 25–30% of the control SYP-1 level following *syp-1* partial RNAi under our experimental conditions. All panels shown are from the same membrane probed with indicated antibodies. Similar results were obtained for three independent experiments. **b**, Representative immunofluorescence images of late pachytene nuclei co-stained for SYP-1 (green), chromosome axis marker HTP-3 (red) and GFP::COSA-1 (blue), showing reduction of SYP-1 fluorescence relative to HTP-3 fluorescence and increase in GFP::COSA-1 foci in *syp-1* partial RNAi nucleus compared to control nucleus. Except for right-most panels, images shown are sum projections through three-dimensional data stacks encompassing whole nuclei. For the first four pairs of control and *syp-1* partial RNAi panels, identical

exposure times and dynamic range settings for image display were used to highlight the reduction in the SYP-1:HTP-3 ratio. In the last two panels, SYP-1 signal levels were adjusted for *syp-1* partial RNAi images to facilitate visualization of the SYP-1 tracts. Because some SCs from the top and bottom halves of the nuclei are superimposed in the full projections encompassing whole nuclei, partial projections showing half nuclei are also provided (right-most images). Scale bar, 2 μ m. **c**, Graphs showing quantification of the reduction in SYP-1 fluorescence relative to HTP-3 fluorescence following *syp-1* partial RNAi. Two different methods for analysing the data (see Methods) yield similar results, indicating that under the *syp-1* partial RNAi conditions used for our experimental analysis, SYP-1 levels are reduced to approximately 30–40% of control levels. Error bars indicate s.d. Numbers of gonads assessed: experiment no. 1: control = 7 gonads, *syp-1* RNAi = 7 gonads; experiment no. 2: control = 15 gonads, *syp-1* RNAi = 6 gonads; experiment no. 3: control = 10 gonads, *syp-1* RNAi = 8 gonads.

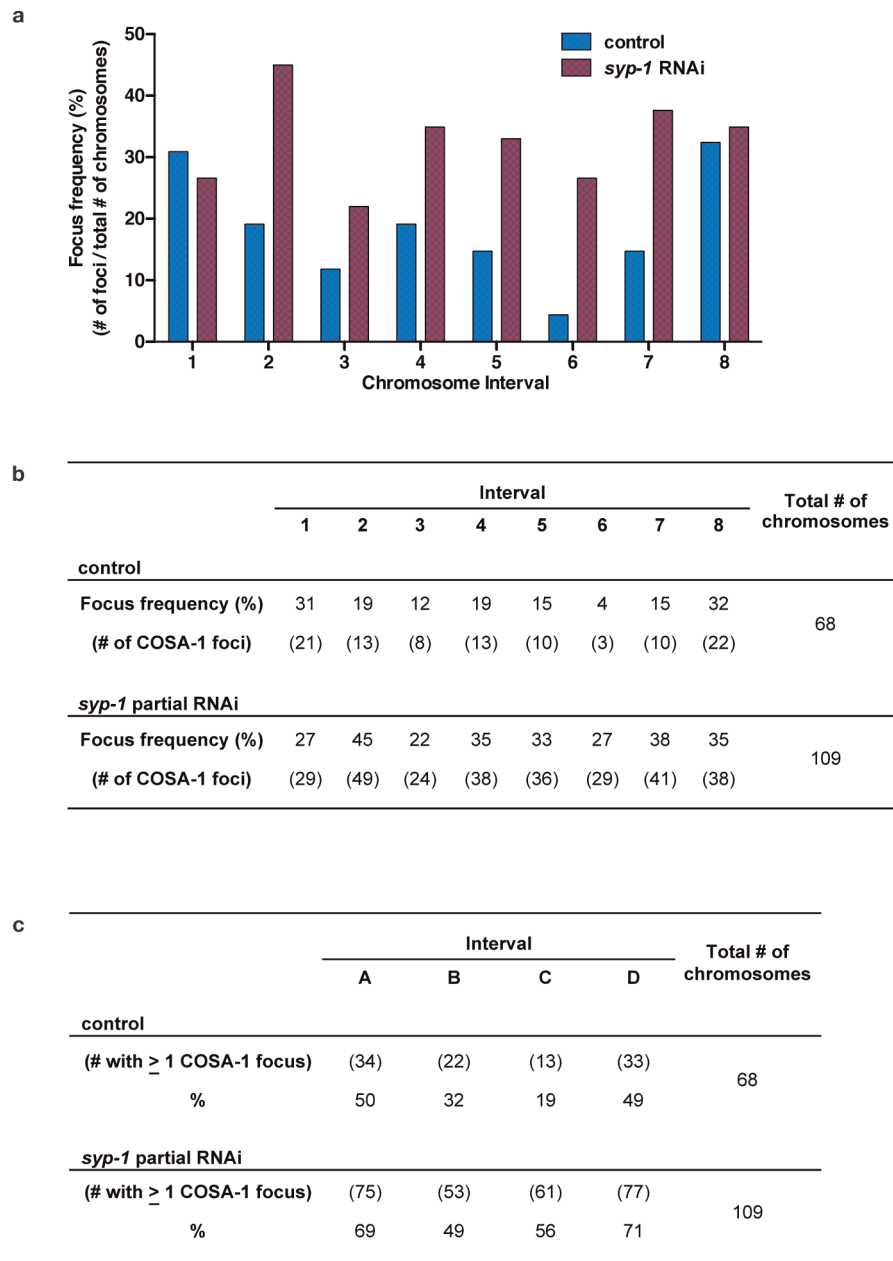


Extended Data Figure 2 | Partial depletion of SYPs increases numbers of COSA-1 foci. Graph depicting the mean numbers of GFP::COSA-1 foci per late pachytene nucleus detected following exposure to *syp-1/F26D2.2*, *syp-2/C24G6.1* or *syp-3/F39H2.4* RNAi or empty vector control²⁸. RNAi and control conditions were identical to those described in Methods, except that worms were dissected for immunofluorescence at 24 h post-L4 stage on RNAi or control plates at 25 °C. Error bars indicate s.d. Control nuclei had an average of six COSA-1 foci per nucleus and a very low standard deviation, indicating operation of the highly robust crossover control system. Partial RNAi treatment for any of the *syp* genes resulted both in a significant increase in the average number of GFP::COSA-1 foci >6 per nucleus (Mann–Whitney, two-tailed $P > 0.0001$ for *syp-1*, *syp-2* and *syp-3* RNAi) and in a much higher s.d., indicating impairment of crossover control. Numbers of nuclei counted were: control, $n = 78$; *syp-1*, $n = 64$; *syp-2*, $n = 129$; *syp-3*, $n = 87$.



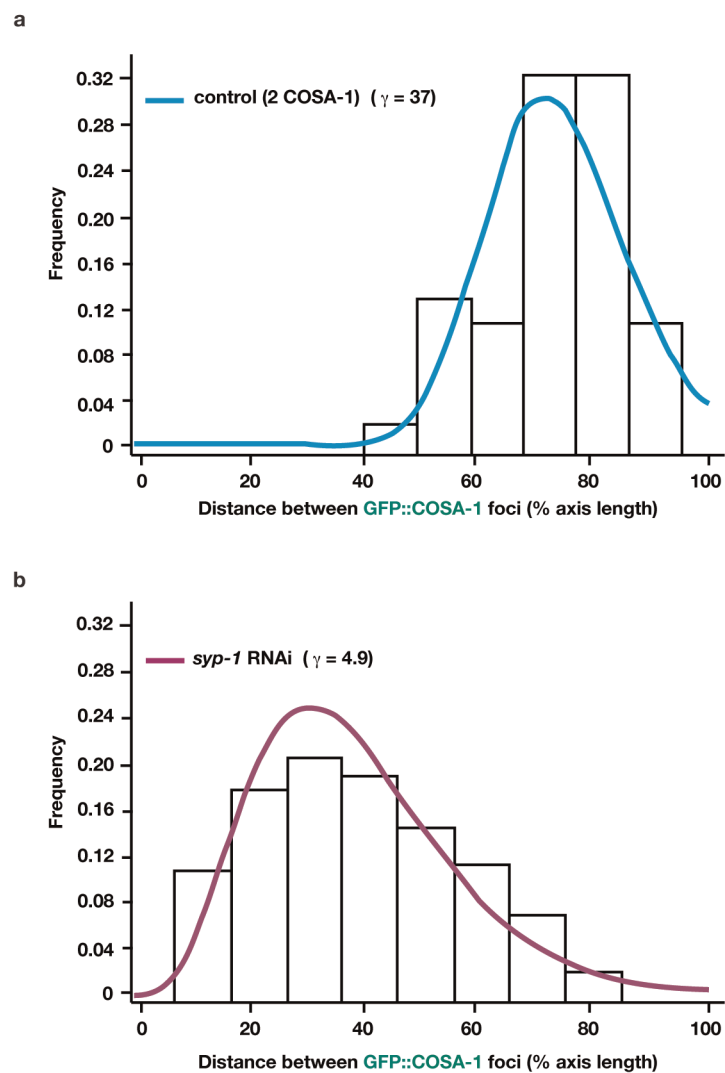
Extended Data Figure 3 | GFP::COSA-1 foci in *syp-1* RNAi nuclei correspond to inter-homologue crossovers. **a**, Quantification of chiasmata on the *mnT12* bivalent in diakinesis-stage oocytes, showing that the incidence of chiasmata corresponds well with the incidence of GFP::COSA-1 foci observed at late pachytene (Fig. 2a). For bivalents with only one or two chiasmata, each individual chiasma was readily scored; bivalents with ≥ 3 chiasmata were pooled into a single category owing to their highly contorted structures, which in some cases made it difficult to discriminate whether 3, 4 or 5 chiasmata were present. In control oocytes, all *mnT12* bivalents had one or two chiasmata. By contrast, 47% of *syp-1* RNAi oocytes had *mnT12* bivalents with ≥ 3 chiasmata, corresponding well with 49% of late pachytene *syp-1* RNAi *mnT12* having ≥ 3 GFP::COSA-1 foci. As formation of each chiasma requires an inter-homologue crossover event, the close correspondence between the numbers of GFP::COSA-1 foci at late pachytene and chiasmata at diakinesis on *syp-1* RNAi *mnT12* bivalents indicates that most, and probably all, GFP::COSA-1 foci in *syp-1* RNAi late pachytene nuclei are marking inter-homologue crossovers. Numbers of diakinesis nuclei scored: control, 86 nuclei from 35 gonads; *syp-1* RNAi, 156 nuclei from 53 gonads. **b**, GFP::COSA-1 foci are not detected on asynapsed chromosome segments. In the regions of the *mnT12* control and *mnT12 syp-1* RNAi germ lines that were imaged for analyses of GFP::COSA-1 foci, we identified a subset of nuclei in which portions of the *mnT12* fusion chromosome pair were asynapsed. These asynapsed segments were found at

comparable frequencies among analysed nuclei from control (8%) and *syp-1* RNAi (12%) germ lines and may represent early stages of desynapsis as cells transition from late pachynema to early diplonema. (Within these nuclei, the asynapsed *mnT12* segments comprised approximately 18% of the total *mnT12* axis length for control and 26% for *syp-1* RNAi.) The *mnT12* bivalents for all control (top left) and *syp-1* RNAi (top right) nuclei in this category are represented in schematic form, with the chromosome axes (HTP-3, red) cartooned to depict both the approximate location and size of the asynapsed segment(s) relative to the total axis length, and the positions of GFP::COSA-1 (green) and HIM-8 (blue) foci. Notably, all GFP::COSA-1 foci on these partially asynapsed chromosome pairs were associated with synapsed segments, located either within a synapsed segment or at the boundary between a synapsed segment and an asynapsed segment; GFP::COSA-1 foci were never found on the asynapsed axis segments. Given the fraction of total axis length that was asynapsed in these *syp-1* RNAi *mnT12* nuclei, the observed restriction of GFP::COSA-1 foci to synapsed segments (where homologues are closely juxtaposed) represents a highly significant (χ^2 test; $P = 0.0002$) departure from the distribution expected if GFP::COSA-1 foci were equally likely to occur on synapsed segments and asynapsed segments (where homologues are separated), consistent with the interpretation that these GFP::COSA-1 foci correspond to inter-homologue recombination events.



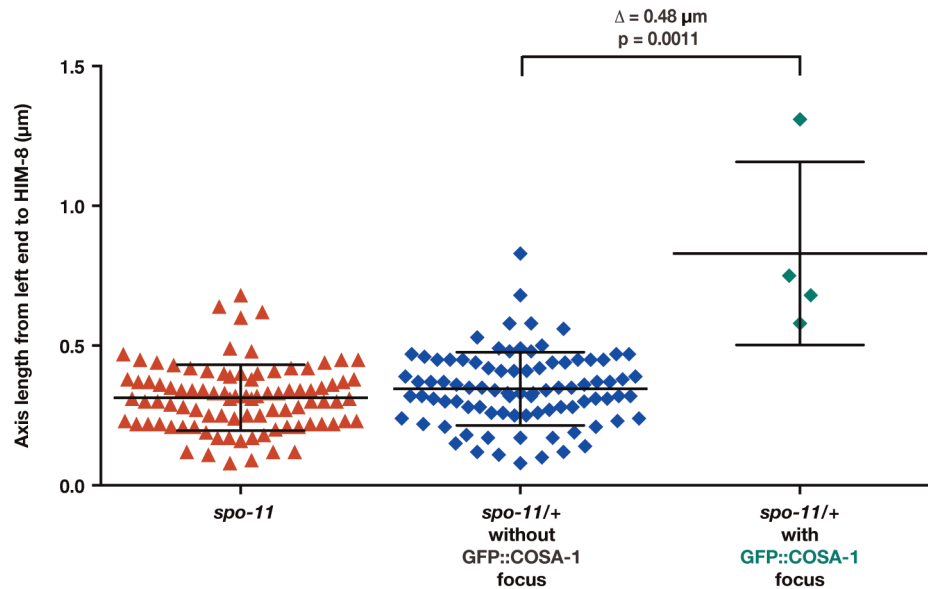
Extended Data Figure 4 | Distribution of GFP::COSA-1 foci among evenly spaced intervals along *mnT12*. **a**, Bar graph for eight-interval analysis of *mnT12* (X;IV) fusion chromosomes, indicating the frequencies of GFP::COSA-1 foci in each interval for control (blue) and *syp-1* partial RNAi (purple) worms. **b**, Table for eight-interval analysis indicating both the focus frequencies and the

numbers of GFP::COSA-1 foci in each interval. **c**, Table for four-interval analysis indicating for each interval both the numbers and the percentages of *mnT12* chromosome pairs with ≥ 1 GFP::COSA-1 focus in that interval (used for interference strength calculations in Fig. 3a and Extended Data Table 1).



Extended Data Figure 5 | Gamma probability distribution modelling of inter-COSA-1 focus distances. **a, b,** Histograms of the distribution of inter-focus distances (reported as percentage of total axis length) for binned control

data ($n = 47$) (**a**) and *syp-1* RNAi data ($n = 183$) (**b**). The best-fit gamma probability distribution curves generated from modelling the binned data sets (Fig. 3a) are overlaid on the histograms.



Extended Data Figure 6 | Association between local axis length and GFP::COSA-1 foci at 20 °C. Scatter plot of length measurements (μm) for the segment of *mnT12* chromosome axis from the left end of *mnT12* to the centre of the HIM-8 focus (as seen in Fig. 4d), for *spo-11/+* nuclei without (blue diamonds) or with (green diamonds) a GFP::COSA-1 focus in this chromosome segment and for the *spo-11* mutant (red triangles), which lacks meiotic DSBs and crossovers. Middle lines indicate mean and error bars

indicate s.d. Mean length measurements for *spo-11* nuclei ($0.31 \mu\text{m}$, $n = 92$) and *spo-11/+* nuclei ($0.35 \mu\text{m}$, $n = 88$) lacking a focus in this chromosome segment were not significantly different from each other (Mann–Whitney, two-tailed $P = 0.062$), whereas both were significantly lower (Mann–Whitney, two-tailed $P = 0.0010$; $P = 0.0011$) than for *spo-11/+* nuclei that had a GFP::COSA-1 focus in this segment ($0.83 \mu\text{m}$, $n = 4$).

Extended Data Table 1 | Four-interval analysis of interference.

		Interval Pair					
		AB	BC	CD	AC	BD	AD
control							
Interference Strength (I)		1	1	1	0.68	0.71	-0.81
p-value (Fisher's exact)*		<0.0001	0.0033	<0.0001	0.0115	<0.0001	<0.0001
<i>syp-1</i> partial RNAi							
Interference Strength (I)		0.53	0.6	0.37	-0.1	-0.1	0.09
p-value (Fisher's exact)*		<0.0001	<0.0001	<0.0001	0.067	0.1448	0.2066

* P values from Fisher's exact tests assessing the probability of obtaining the observed data set assuming independent behaviour of the two intervals.

Visualizing virus assembly intermediates inside marine cyanobacteria

Wei Dai¹, Caroline Fu¹, Desislava Raytcheva^{2,3}, John Flanagan^{1†}, Htet A. Khant¹, Xiangnan Liu¹, Ryan H. Rochat^{1,4}, Cameron Haase–Pettingell², Jacqueline Piret³, Steve J. Ludtke^{1,4}, Kuniaki Nagayama⁵, Michael F. Schmid^{1,4}, Jonathan A. King² & Wah Chiu^{1,4}

Cyanobacteria are photosynthetic organisms responsible for ~25% of organic carbon fixation on the Earth. These bacteria began to convert solar energy and carbon dioxide into bioenergy and oxygen more than two billion years ago. Cyanophages, which infect these bacteria, have an important role in regulating the marine ecosystem by controlling cyanobacteria community organization and mediating lateral gene transfer. Here we visualize the maturation process of cyanophage Syn5 inside its host cell, *Synechococcus*, using Zernike phase contrast electron cryo-tomography (cryoET)^{1,2}. This imaging modality yields dramatic enhancement of image contrast over conventional cryoET and thus facilitates the direct identification of subcellular components, including thylakoid membranes, carboxysomes and polyribosomes, as well as phages, inside the congested cytosol of the infected cell. By correlating the structural features and relative abundance of viral progeny within cells at different stages of infection, we identify distinct Syn5 assembly intermediates. Our results indicate that the procapsid releases scaffolding proteins and expands its volume at an early stage of genome packaging. Later in the assembly process, we detected full particles with a tail either with or without an additional horn. The morphogenetic pathway we describe here is highly conserved and was probably established long before that of double-stranded DNA viruses infecting more complex organisms.

Cyanobacteria, and presumably their cyanophages, predate the emergence of enteric bacteria and mammalian viruses by ~2.7 billion years. Cyanophages containing double-stranded (ds)DNA infect a wide range of photosynthetic cyanobacteria. A key question in the assembly of dsDNA viruses is the coordination between protein shell assembly and genome packaging. Here we use a relatively new electron cryo-microscopy approach to follow the maturation process of wild-type cyanophage Syn5 as it occurs inside its host, *Synechococcus* sp. WH8109^{3,4}.

We preserved the native structure of the phage-infected cells by plunge-freezing⁵ and maintaining them at liquid nitrogen temperature during imaging. The frozen, hydrated cells were imaged in an electron microscope equipped with a Zernike phase plate, a thin carbon film with a central hole, placed in the back focal plane of the objective lens^{1,2}. It shifts the phase of the scattered electrons by $\pi/2$, analogous to an optical phase contrast microscope. This significantly enhances the low-frequency information, allowing for in-focus, high-contrast imaging^{6–8} (Extended Data Fig. 1). Consequently, low-contrast features difficult to detect in conventional cryoET images can be more readily identified.

WH8109 cells were imaged before infection and 65–70 min after infection. Even at this late infection time, some cells seemed to be newly infected. We reconstructed 58 Zernike phase contrast (ZPC) tomograms of WH8109 cells (Figs 1a and 2, Supplementary Videos 1–4 and Methods). The cells range from 0.7 to 1.0 μm in diameter. Although the cell envelope and thylakoid membrane (Fig. 1a, b) are roughly concentric,

the thylakoid membrane does not fully enclose the inner compartment of the cell, nor does it seem to directly interact with the cell membrane. This differs from the organization seen in other cyanobacteria^{9,10}. Cyanobacteria also contain carboxysomes, polyhedral compartments encapsulating enzymes for carbon fixation^{11,12}. Each WH8109 cell has, on average, four or five carboxysomes, with diameters ranging from 920 to 1,160 Å (Fig. 1c). Ribosomes are abundant and widespread, forming numerous intracellular patches that contain polyribosomes (Fig. 1d).

Cyanophage Syn5 that infects WH8109 cells is a short-tailed icosahedral phage with a unique horn appendage at the vertex opposite to the tail¹³ (Extended Data Fig. 2). Initial segmentation of our tomograms

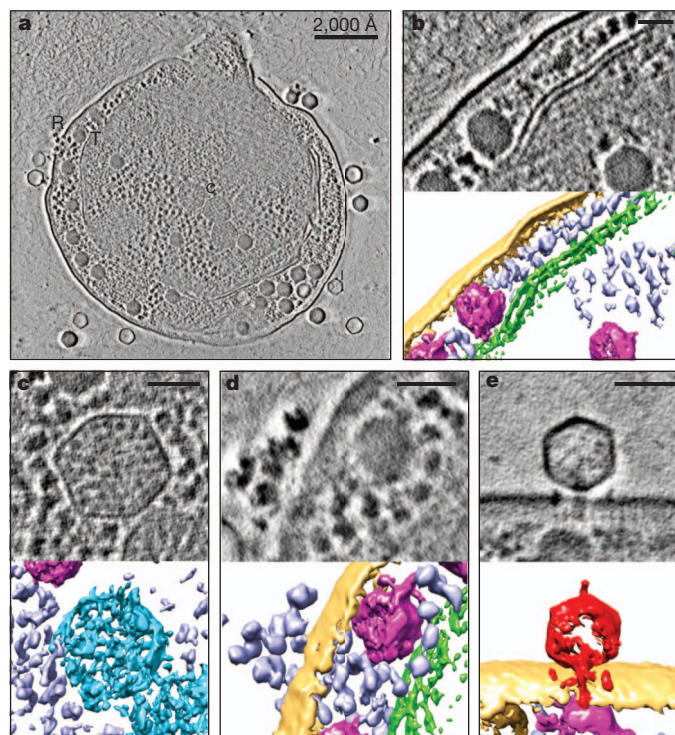


Figure 1 | ZPC cryoET enables direct recognition of cellular components of the Syn5-infected WH8109 cells. **a**, Section view of a Syn5-infected cell at a late stage of infection with components labelled, including ribosomes (R), thylakoid membranes (T), carboxysomes (C) and infecting phages (I). **b–e**, Section and three-dimensional annotated view of cellular components in **a** are shown. **b**, Thylakoid membrane (green). **c**, Carboxysome (blue). **d**, Ribosome (purple). **e**, An infecting Syn5 phage (red) positioned normal to the surface of the infected cell. Yellow, cell envelope; magenta, phage progeny. Scale bars, 500 Å (**b**, **c**) and 600 Å (**d**, **e**).

¹National Center for Macromolecular Imaging, Verna and Marrs Mclean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ²Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA. ⁴Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA. ⁵National Institute for Physiological Sciences, National Institutes of Natural Sciences, 5-1 Higashi-yama, Myodaiji, Okazaki 444-8787, Japan. [†]Present address: FEI, 5350 Dawson Creek Drive, Hillsboro, Oregon 97124, USA.

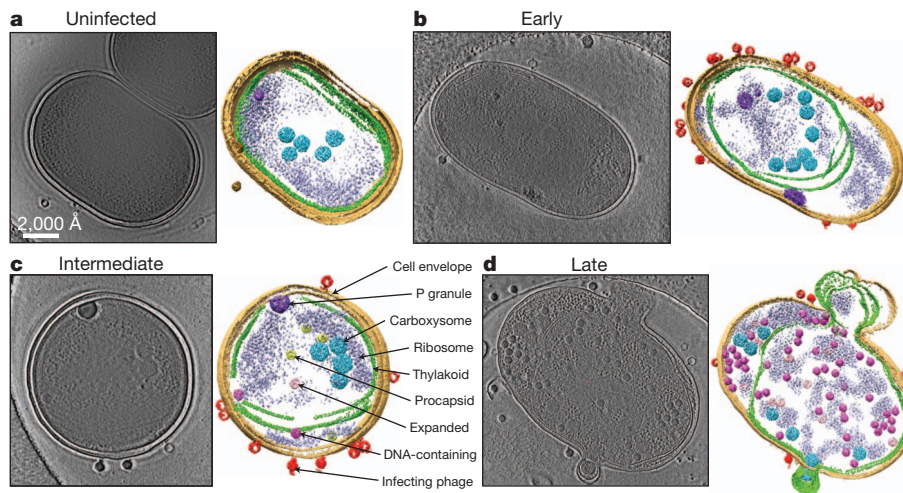


Figure 2 | ZPC cryoET of WH8109 cells before and after infection with Syn5 phage. **a–d**, Section (left) and annotated (right) views of an uninfected cell (**a**) and infected cells at early (**b**), intermediate (**c**) and late (**d**) stages of infection. Sections shown are 54 Å slabs taken from the middle of the

tomograms. Cellular components and phages are coloured and labelled in the annotated view in **c**. Phage progeny can be separated into three types on the basis of size, shape and internal density: procapsid, yellow; expanded capsid, pink; and DNA-containing capsid, magenta.

of infected cells identified Syn5 particles on the cell surface, floating in the extracellular medium, and Syn5 progeny inside the cell. Multiple full and empty phage particles are seen attached to the cell surface. Injection of viral DNA occurs at multiple sites on the bacterial envelope and does not seem to be a coordinated process. Figure 1e shows a tubular density extending from the phage tail through the periplasm to the cytoplasm (Supplementary Video 4), similar to observations in other phage-infected bacteria^{14,15}. As infection progresses, increasing numbers of Syn5 phage progeny are observed inside the cells. Late in infection, the cell membrane deforms and ruptures, releasing the phage progeny (Fig. 2).

We extracted 470 subvolumes of intracellular Syn5-like particles and classified them into three morphological types on the basis of their shape, size and internal density. The particles were then subjected to template-free alignment and classification^{16,17} to obtain averages for each type (Methods). The resolutions of the averages range between 70 to 50 Å. This level of resolution is sufficient to support our structural interpretations.

The most recognizable type of intracellular capsid appears similar in size (~660 Å in diameter) and shape to the mature Syn5 phage¹³ (Fig. 3a–c). Particles of this type represent the largest population, and are especially abundant in cells at later stages of infection. They have an icosahedral capsid shell with significant internal density attributable to DNA, and are hereafter referred to as DNA-containing capsids. In contrast to the homogenous population of isolated mature phage, we observed three subtypes of these particles inside infected cells, differing at two opposing vertices. They represent particles with (1) a bulky tail and a slim horn appendage on opposing vertices, as in the mature phage (Fig. 3a); (2) a tail at one vertex only (Fig. 3b); and (3) no detectable density protruding from any vertex (Fig. 3c). The averages of the first two subtypes (Fig. 3a, b) show a tail hub of length 190 Å; tail fibres are not usually resolved. This could be due to incomplete tail assembly at intermediate stages, inherent flexibility of the tail fibres and/or interference from neighbouring intracellular densities. Our recognition of these three subtypes reveals that the assembly of the tail hub follows DNA encapsulation, but precedes the addition of the horn.

The second phage progeny type consists of spherical particles that are ~10% smaller than the mature Syn5 (refs 13, 18). They have a diameter of 590 Å and shell thickness of ~50 Å (Fig. 3e). In the subvolume average, there is a density extending inward at one location of the shell, which we interpret to be the Syn5 portal. However, we could not locate the icosahedral symmetry axes in these particles because of their spherical shape, and thus cannot definitively assign the inward

pointing portal density to an icosahedral vertex. This putative portal is connected to an inner spherical density (~220 Å in diameter) that resembles the scaffolding proteins seen in procapsids of other dsDNA phages^{19,20}. Given these characteristics, we conclude that this particle type represents Syn5 procapsids¹⁸.

The third particle type is characterized by an angular shape (Fig. 3d) and has not previously been reported in *in vitro* structural studies. Fifty-three subvolumes comprise this group. Thirty-seven are quantitatively confirmed to have icosahedral symmetry (Methods). Model-free alignment of the 37 particles produced an average with an icosahedral shell and a ~200 Å internally protruding density at one of its 12 vertices. The size of the average (~660 Å diameter, 40 Å shell thickness) matches that of mature Syn5 (refs 13, 18). The internally protruding density could correspond to the full-length portal protein complex, possibly with additional proteins and/or DNA. Although the raw individual particles display internal density of variable contrast, shape, size and distribution, their average appears empty (except for the portal) because the locations of these densities with respect to the capsid and the portal are not uniform and are thus averaged out. In closely examining this subvolume average, we noticed that a flat protruding density (~40 Å in length) outside the capsid shell is present at the portal vertex. This density cannot correspond to the tail hub, which is assembled after the DNA is fully encapsulated. Although our resolution is insufficient to positively identify the density as being the terminase, this density occurs at the position and stage of assembly expected for the terminase²¹.

This third particle type may correspond to either an abortive particle or a functional intermediate between the procapsid and DNA-containing capsid types. We carried out the following analysis to resolve this ambiguity. Phage attachment to the outer cell membrane and DNA injection into the host cells are not synchronized under our experimental conditions. Therefore, multiple stages of infection and phage assembly are present in different cells at 65 min after phage infection. To investigate the order in which the above phage types occur before maturation, we used the number of DNA-containing particles (type 1) per cell as a proxy for the progression of productive infection (Fig. 3f)—the greater the number of DNA-containing capsids inside a cell, the farther the infection has proceeded. As infection progresses, the total number of phage progeny observed increases from 2 to 84 per cell. In cells at early infection stages, procapsids and expanded capsids are seen before DNA-containing capsids appear (Fig. 3f, inset). This indicates that procapsids and expanded progeny are assembled before the DNA-containing capsids. With progression of infection, the numbers of procapsid and

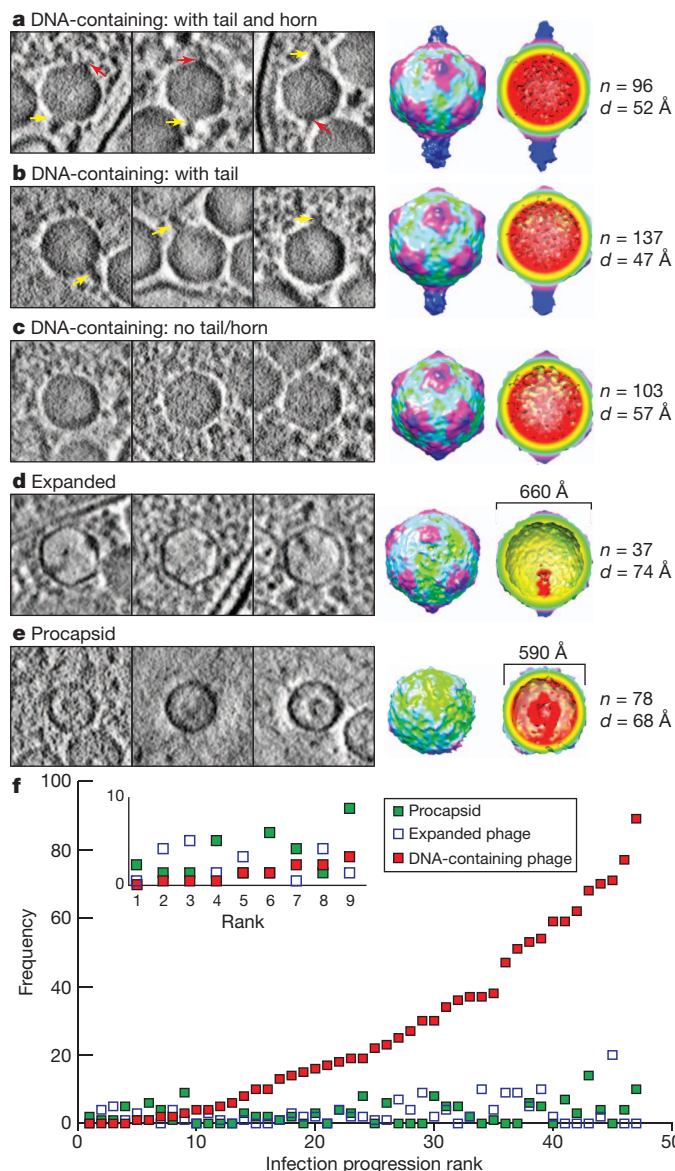


Figure 3 | Phage progeny average maps reveal diverse assembly intermediates during phage assembly. **a–e**, Phage progeny classified as DNA-containing capsid (**a–c**), expanded capsid (**d**) and procapsid (**e**). The left three panels show 54 Å slabs containing representative particles. Yellow arrows, tails; red arrows, horns. Right panels show the averaged maps, with the number of subvolumes (n) and resolutions (d). **f**, Number of the three types of phage progeny with progression of infection. Forty-seven tomograms of intact cells were ranked by the number of DNA-containing capsids. The numbers of the phage progeny in each cell were plotted against the cell tomogram ranking. The inset shows an expansion of the first ten ranks.

expanded capsids increase only slightly. This slow rate of increase suggests that these species are short-lived, and progeny exit these states at almost the same rate as they enter them throughout the infection process. The lack of accumulation of procapsid and expanded capsids is consistent with previous biochemical experiments¹⁸, and supports the notion that the expanded particles are assembly intermediates after the capsid shell has expanded and acquired icosahedral angularity.

It was not previously known whether shell expansion and DNA encapsulation occur sequentially or simultaneously^{22,23}. Our identification of the expanded intermediates (Fig. 3d) reveals that, in Syn5 and probably some other phages²⁴, the conformational changes of the capsid, the expansion of the shell, and acquisition of angularity, are completed before the full length of viral DNA is packaged (Fig. 4). Those expanded capsids without icosahedral symmetry (16 out of 53, and not

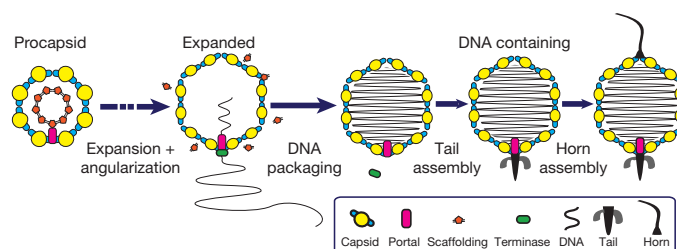


Figure 4 | Phage assembly model revealed by ZPC cryoET. The phage's pathway to maturation starts with the assembly of the precursor procapsid. Initial insertion of DNA into the capsid induces capsid expansion and angularity. The reorganization of the capsid shell culminates before DNA is fully packaged into it. The tail is then added to the portal vertex after DNA packaging. As the final capsid assembly step, the horn is attached to the vertex opposite to the tail.

included in the average shown in Fig. 3d) may be in the midst of transformation from procapsid to the expanded capsid upon DNA entry.

The intracellular assembly of many dsDNA phages and viruses—including adenoviruses and herpesvirus—proceeds through assembly of a precursor procapsid shell, containing scaffolding proteins and a cyclic portal complex defining a unique vertex. Our results show that Syn5 shares the same procapsid-forming pathway as enteric bacteriophages and eukaryotic viruses (Fig. 4). Because cyanobacteria preceded enteric bacteria in evolution, it is reasonable to propose that enteric bacteriophages might have inherited this assembly pathway from cyanophages.

We examined the impact of phage infection on cell physiology by evaluating subcellular components throughout infection. The number of carboxysomes, which often reflects the cellular metabolism level, remained invariant (Extended Data Fig. 3). This observation suggests that phage production does not profoundly perturb the host cell's metabolism until lysis.

Our study demonstrates the first application of ZPC cryoET to examine cellular processes without labelling or sectioning. Post-tomographic analyses allowed us to mine the rich trove of spatial and temporal information conveyed by the complex biological process of phage infection and maturation *in situ*. The value of our imaging approach lies in its ability to study the ancient process of phage assembly in its natural intracellular environment at nanometre resolution, offering the potential to characterize cyanobacterial strains modified for a wide range of applications, including bioenergy development²⁵.

METHODS SUMMARY

Synechococcus sp. strain WH8109 cells were grown in artificial sea water¹³. Cells at exponential phase were infected with Syn5 phage at a multiplicity of infection (m.o.i.) of 5. At 65–75 min after infection, the cells were collected for cryo-specimen preparation. Tilt series of frozen, hydrated cells were collected in a JEM2200FS electron microscope (JEOL) operated at 200 kV and specially equipped with a $\pi/2$ thin carbon film Zernike phase plate^{2,6,8}. Low-dose tilt series were manually collected. IMOD²⁶ was used to align tilt series and reconstruct tomograms.

Subvolumes of Syn5 progeny phages were extracted from tomograms of infected cells and visually classified into three types based on size, shape and intra-capsid density. For each type, a symmetry-searching algorithm^{27,28} was used to determine whether the particles possessed icosahedral symmetry. If symmetry was confirmed, one of the 12 vertices and a twofold axis of each particle were aligned along the z and y axes, respectively. Expanded and DNA-containing capsids that had detectable icosahedral symmetry were then subjected to further template-free classification and alignment^{16,17,29} to bring the portal vertex into proper register across particles. The contrast of the subvolumes extracted from our ZPC tomograms was high enough to classify them both visually and quantitatively. For procapsids, the symmetry-searching program failed to identify icosahedral symmetry, so they were aligned using hierarchical ascendant classification^{16,17,29} to obtain their final average. The resolution of each of the particle averages was estimated by computing the 0.143 Fourier shell correlation criterion³⁰ from two independent subsets of averages of each type and subtype particle subvolumes.

Segmentation and annotation of the cell tomograms were done in Avizo (Visualization Sciences Group, FEI).

General linear modelling was performed to correlate carboxysome number with progression of infection using PROC GLM.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 May; accepted 27 August 2013.

Published online 9 October 2013.

- Danev, R. & Nagayama, K. Phase plates for transmission electron microscopy. *Methods Enzymol.* **481**, 343–369 (2010).
- Murata, K. *et al.* Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions. *Structure* **18**, 903–912 (2010).
- Fuller, N. J. *et al.* Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the red sea. *Appl. Environ. Microbiol.* **69**, 2430–2443 (2003).
- Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* **68**, 1180–1191 (2002).
- Taylor, K. A. & Glaeser, R. M. Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. *J. Struct. Biol.* **163**, 214–223 (2008).
- Danev, R., Glaeser, R. M. & Nagayama, K. Practical factors affecting the performance of a thin-film phase plate for transmission electron microscopy. *Ultramicroscopy* **109**, 312–325 (2009).
- Marko, M., Leith, A., Hsieh, C. & Danev, R. Retrofit implementation of Zernike phase plate imaging for cryo-TEM. *J. Struct. Biol.* **174**, 400–412 (2011).
- Rochat, R. H. *et al.* Seeing the portal in herpes simplex virus type 1 B capsids. *J. Virol.* **85**, 1871–1874 (2011).
- Liberton, M., Austin, J. R. II, Berg, R. H. & Pakrasi, H. B. Unique thylakoid membrane architecture of a unicellular N₂-fixing cyanobacterium revealed by electron tomography. *Plant Physiol.* **155**, 1656–1666 (2011).
- Ting, C. S., Hsieh, C., Sundararaman, S., Mannella, C. & Marko, M. Cryo-electron tomography reveals the comparative three-dimensional architecture of *Prochlorococcus*, a globally important marine cyanobacterium. *J. Bacteriol.* **189**, 4485–4493 (2007).
- Iancu, C. V. *et al.* Organization, structure, and assembly of α -carboxysomes determined by electron cryotomography of intact cells. *J. Mol. Biol.* **396**, 105–117 (2010).
- Schmid, M. F. *et al.* Structure of *Halothiobacillus neapolitanus* carboxysomes by cryo-electron tomography. *J. Mol. Biol.* **364**, 526–535 (2006).
- Pope, W. H. *et al.* Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a “horned” bacteriophage of marine *synechococcus*. *J. Mol. Biol.* **368**, 966–981 (2007).
- Chang, J. T. *et al.* Visualizing the structural changes of bacteriophage Epsilon15 and its *Salmonella* host during infection. *J. Mol. Biol.* **402**, 731–740 (2010).
- Hu, B., Margolin, W., Molineux, I. J. & Liu, J. The bacteriophage τ 7 virion undergoes extensive structural remodeling during infection. *Science* **339**, 576–579 (2013).
- Schmid, M. F. Single-particle electron cryotomography (cryoET). *Adv. Protein Chem. Struct. Biol.* **82**, 37–65 (2011).
- Schmid, M. F. & Booth, C. R. Methods for aligning and for averaging 3D volumes with missing data. *J. Struct. Biol.* **161**, 243–248 (2008).
- Raytcheva, D. A., Haase-Pettingell, C., Piret, J. M. & King, J. A. Intracellular assembly of cyanophage Syn5 proceeds through a scaffold-containing procapsid. *J. Virol.* **85**, 2406–2415 (2011).
- Chen, D. H. *et al.* Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc. Natl Acad. Sci. USA* **108**, 1355–1360 (2011).
- Prevelige, P. E., Thomas, D. & King, J. Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells *in vitro*. *J. Mol. Biol.* **202**, 743–757 (1988).
- Hegde, S., Padilla-Sanchez, V., Draper, B. & Rao, V. B. Portal-large terminase interactions of the bacteriophage T4 DNA packaging machine implicate a molecular lever mechanism for coupling ATPase to DNA translocation. *J. Virol.* **86**, 4046–4057 (2012).
- Aksyuk, A. A. & Rossmann, M. G. Bacteriophage assembly. *Viruses* **3**, 172–203 (2011).
- Hendrix, R. W. Bacteriophage assembly. *Nature* **277**, 172–173 (1979).
- Fuller, D. N. *et al.* Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.* **373**, 1113–1122 (2007).
- Machado, I. M. & Atsumi, S. Cyanobacterial biofuel production. *J. Biotechnol.* **162**, 50–56 (2012).
- Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
- Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
- Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
- Schmid, M. F. *et al.* A tail-like assembly at the portal vertex in intact herpes simplex type-1 virions. *PLoS Pathog.* **8**, e1002961 (2012).
- Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by grants from the Robert Welch Foundation (Q1242) and National Institutes of Health (P41GM123832 to W.C.; AI0175208 and PN2EY016525 to W.C. and J.A.K.; GM080139 to S.J.L.; T15LM007093 through the Gulf Coast Consortia to W.D. and R.H.R.; T32GM007330 through the MSTP to R.H.R.). We thank J. G. Galaz-Montoya and R. N. Irobalieva for editing of the manuscript.

Author Contributions W.D., D.R. and C.F. prepared the samples and conducted the infection experiments under the advice of C.H.-P. and J.P. W.D. collected the image data and reconstructed the tomograms; C.F. and H.A.K. established the Zernike phase plate imaging conditions in the microscope; K.N. provided the phase plates for imaging; R.H.R. performed the statistical analysis. W.D. and M.F.S. developed the imaging processing methods and solved the structures of the phage assembly intermediates; W.D. and X.L. refined the structures; J.F. and S.J.L. developed the symmetry-search algorithm for subvolume alignment; W.D., M.F.S., J.A.K. and W.C. interpreted the structures and wrote the manuscript.

Author Information The averaged density maps of the procapsid, expanded capsid and the DNA-containing capsid have been deposited in the EBI under accession codes EMD-5742, EMD-5743, EMD-5744, EMD-5745 and EMD-5746, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.C. (wah@bcm.edu).

METHODS

Host growth and infection. The Syn5 host *Synechococcus* sp. strain WH8109 cells were grown in artificial sea water (ASW)¹³ with continuous aeration in gas dispersion bottles. To make the ASW medium, nine salts (428 mM NaCl, 9.8 mM MgCl₂·6H₂O, 6.7 mM KCl, 17.8 mM NaNO₃, 14.2 mM MgSO₄, 3.4 mM CaCl₂·2H₂O, 0.22 mM K₂HPO₄·3H₂O, 5.9 mM NaHCO₃, 9.1 mM Tris) were dissolved in MilliQ water, then pH was adjusted to 8.0 with HCl. After autoclaving and cooling down to room temperature, a trace solution was added to the following final concentrations: 0.77 μM ZnSO₄·7H₂O, 7.0 μM MnCl₂·4H₂O, 0.14 μM CoCl₂·6H₂O, 30 μM Na₂MoO₄·2H₂O, 30 μM citrate, 5 μM ferric citrate. Also, a supplementary salt solution was added to a final concentration of 100 μM Na₂CO₃ and 15 μM Na₂EDTA·2H₂O.

For the infection experiments, cells at exponential phase were infected with Syn5 phage at a multiplicity of infection (m.o.i.) of 5. The cells then were kept in a controlled water bath at 28 °C under continuous cool white fluorescent light at an intensity of 1,200 lx, and a shaking speed of 110 r.p.m. At 65–75 min after infection, cells were centrifuged at 8,500g for 5 min. The cell pellet was gently re-suspended in fresh ASW medium and concentrated 100-fold.

Tomographic tilt series acquisition and reconstruction. For cryoET, Syn5-infected WH8109 cell sample was first mixed with 100 Å gold fiducial markers to facilitate alignment in data processing. An aliquot of 3.5 μl sample was applied to 2.0/1.0 μm Quantifoil holey grids (Quantifoil) and plunge frozen using a Vitrobot (FEI). The frozen, hydrated samples were transferred to a Gatan cryo-holder (Gatan, Inc.) and kept at −170 °C throughout the imaging session in a JEM2200FS electron microscope (JEOL). This electron microscope has an in-column energy filter and the slit was set to 15 eV. In addition, the microscope was equipped with an airlock system to allow insertion of $\pi/2$ Zernike phase plate made of a thin carbon film with a 0.7 μm central hole^{2,6,8}. The illumination setting used was: spot size 1; condenser aperture = 70 μm; objective aperture = 60 μm.

Tilt series of frozen, hydrated samples were collected manually with an electron energy of 200 kV under low-dose conditions on a Gatan 4k × 4k CCD camera (Gatan, Inc.) at ×25,000 microscope magnification. The sampling of the data was calibrated to be 4.52 Å per pixel. Typically, a tilt series ranged from −60° to 60° at 3° step increment. The accumulated exposure for each tilt series was 40–50 electrons Å^{−2}. We set the defocus to near zero for imaging a 0° tilt specimen as judged by the computed fast Fourier transform (FFT) of a region next to the specimen of interest, and then used pre-calibrated defocus change for subsequent tilts within a tilt series. We could not afford to use computer-aided focusing for every tilt, because we wanted to minimize the damage to the phase plate from electron exposure. However, we recorded a moderate dose image during our tilt series every 5–10 images to monitor the defocus setting, and to determine the charging status of the phase plate. In general, the defocus values were consistent within 1 μm. The manner and extent of change observed in the FFT varies depending on the initial quality of the phase plate and the cumulative electron exposure to the phase plate. In certain cases, a ring similar to a contrast transfer function (CTF) zero appears at low spatial frequency, as previously shown⁷. In other cases, an FFT of the image can display different patterns indicative of charging. Once a phase plate was found to suffer from charging, we replaced it with a new one. Of note, there were many times that we had to stop image acquisition in the middle of a tilt series because there was no good phase plate replacement available.

IMOD²⁶ was used to align tilt series and reconstruct tomograms. Because the images of tilted specimens were taken with low dose, it was impossible to detect the CTF rings. However, with resolution limited to ~50–70 Å, we can tolerate defocus as high as 5 μm without a need for CTF phase flipping. Our targeted defocus values were substantially less than this, so CTF correction was not necessary.

Subvolume alignment and averaging. Subvolumes of progeny phages, found inside infected WH8109 cells, were extracted from 22 good cell tomograms. Initial classification of the procapsid, the expanded capsid and the DNA-containing capsid was done by visual inspection based on the size and shape of the particles. The classified particles were then subjected to symmetry search^{27,28} (e2symsearch.py from EMAN2²⁸, described below), alignment, classification and averaging.

For DNA-containing capsids (type 1 particles) we used the symmetry-searching algorithm to align particles to the icosahedral symmetry axes. Then the particles were subject to a model-free all-vs-all alignment scheme (incorporating hierarchical ascendant classification^{16,17,29}), constrained to search only the 12 vertices to obtain an initial model, in which the special vertex in each particle (containing the portal) was registered to the positive *z* axis. The high contrast of the ZPC cryoET subvolumes allowed us to further partition the particles manually into three subtypes: with neither tail nor horn, with tail only, and with tail and horn on opposite vertices.

The expanded capsid particles (type 3) have an apparent angular shape. We used the symmetry-searching algorithm to determine whether the particles had icosahedral symmetry. Those with identified icosahedral symmetry (37 out of 53) were subjected to the alignment scheme described above to bring the portal vertex in each particle into proper register²⁹. The final average shows few missing wedge artefacts. In addition, the strong icosahedral symmetry of the capsid shell validates our symmetry-free alignment and averaging procedure.

The procapsid particles (type 2 subvolume) are smooth and pseudo-spherical, without apparent icosahedral symmetry as determined by the symmetry search algorithm. We thus used the model-free alignment scheme searching the entire rotational space to obtain an initial model from a subset of particles. We iteratively refined the alignment of the entire data set using this initial model until no further improvement was observed. The final average lacked detectable icosahedral symmetry.

Symmetry searching algorithm. To align the Syn5 particles to their symmetry axis, we developed a new algorithm: first the algorithm takes a particle in its current orientation and applies icosahedral symmetry to it. If the orientation of the raw particle is aligned near the conventional icosahedral symmetry axes as defined in EMAN²⁷, the symmetrized particle's icosahedral features will be enhanced. If the particle is not so aligned, the symmetrization will smear out the particle into a ball. Next the algorithm computes the normalized cross-correlation between the symmetrized and unsymmetrized particles^{16,17,29}, accounting for the missing wedge. A GSL multidimensional simplex minimizer (<http://www.gnu.org/software/gsl/>) varies the three Euler angles and three translation coordinates of the raw particle, applies icosahedral symmetry to the particle in this new orientation, and computes a new cross-correlation score. The GSL minimizer continues to generate new particle orientations by going downhill until the (negative of the) cross-correlation decreases by less than 0.01.

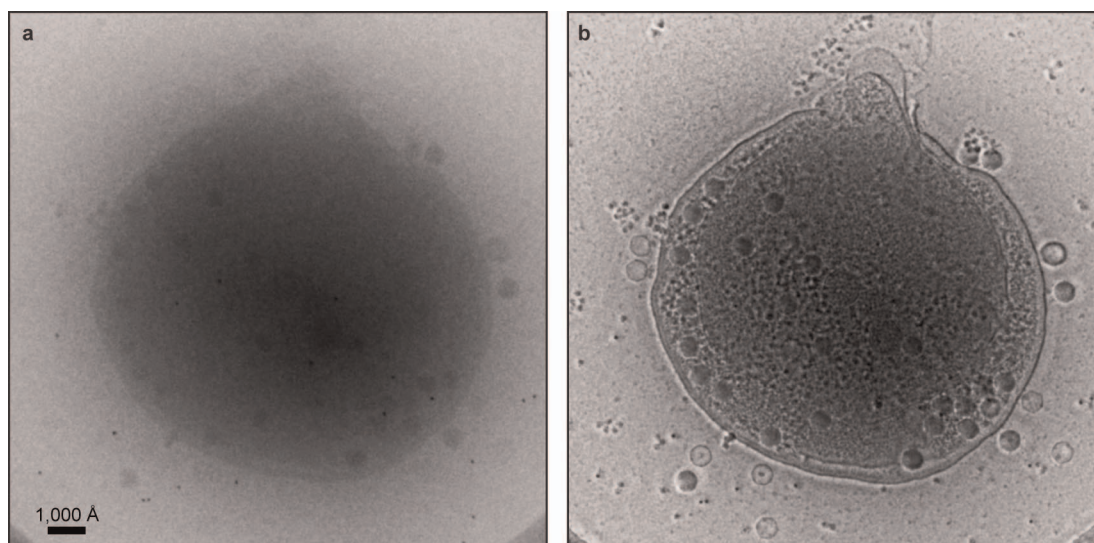
The simplex minimizer does not guarantee a global minimum. We thus repeated the above process '*n*' times, each time applying a random rotation to the same particle as generally used in a Monte Carlo optimization algorithm. In this process, we will find the global minimum of the cross-correlation if '*n*' is sufficiently big. To speed up the search, we centred the particles before performing the cross-correlation computation. In our experience *n* = 10 is sufficient to determine if the particle has icosahedral symmetry. The judgment of whether the particle has icosahedral symmetry is based not only on the score but also a visual inspection, that is, the symmetrized particle should have distinct vertices and the correct capsid surface features.

Resolution estimates of the subvolume averages. There is not yet a widely accepted and rigorous standard for resolution estimates of subvolume averages; however, the gold-standard Fourier shell correlation (FSC) methodology^{30,31} used in single particle analysis can be adapted to produce good estimates. Each of the subvolumes for the five phage assembly intermediates were split into two subsets, to compute two independent averages for each, then the FSC was computed between the two averages. The resolution for the combined average of each phage assembly intermediate is then measured using an FSC threshold of 0.143 (refs 30, 32).

Cut-on frequency correction. The effect on low-resolution contrast is primarily due to the cut-on frequencies imposed by the central hole in the Zernike phase plate. The Fourier components of these images below 1/300 Å^{−1} would be lost and between 1/300–1/20 Å^{−1} would be highly enhanced. We undertook the following steps to re-scale the Fourier components of the map. We computed a 1D structure factor derived from a published single particle reconstruction of Syn5¹³ from conventional cryo-electron microscopy (cryo-EM). This '1D structure factor' is simply the rotationally averaged power spectrum of the particle structure. It is applied to our subvolume averages such that their 1D structure factors match this 'known' curve. As a rotationally symmetric linear filter, this does not impose any new features on the structure. We used the EMAN1²⁷ proc3d command to process the averaged subvolumes with the options of 'apix=9.04 setsf=syn5-structure-factor lp=40'. This corrects the amplitudes, and then applies a 40 Å low-pass filter to suppress high-resolution noise. Note that no correction was attempted for subvolume averages of procapsid and expanded capsid because an appropriate structure factor was not available.

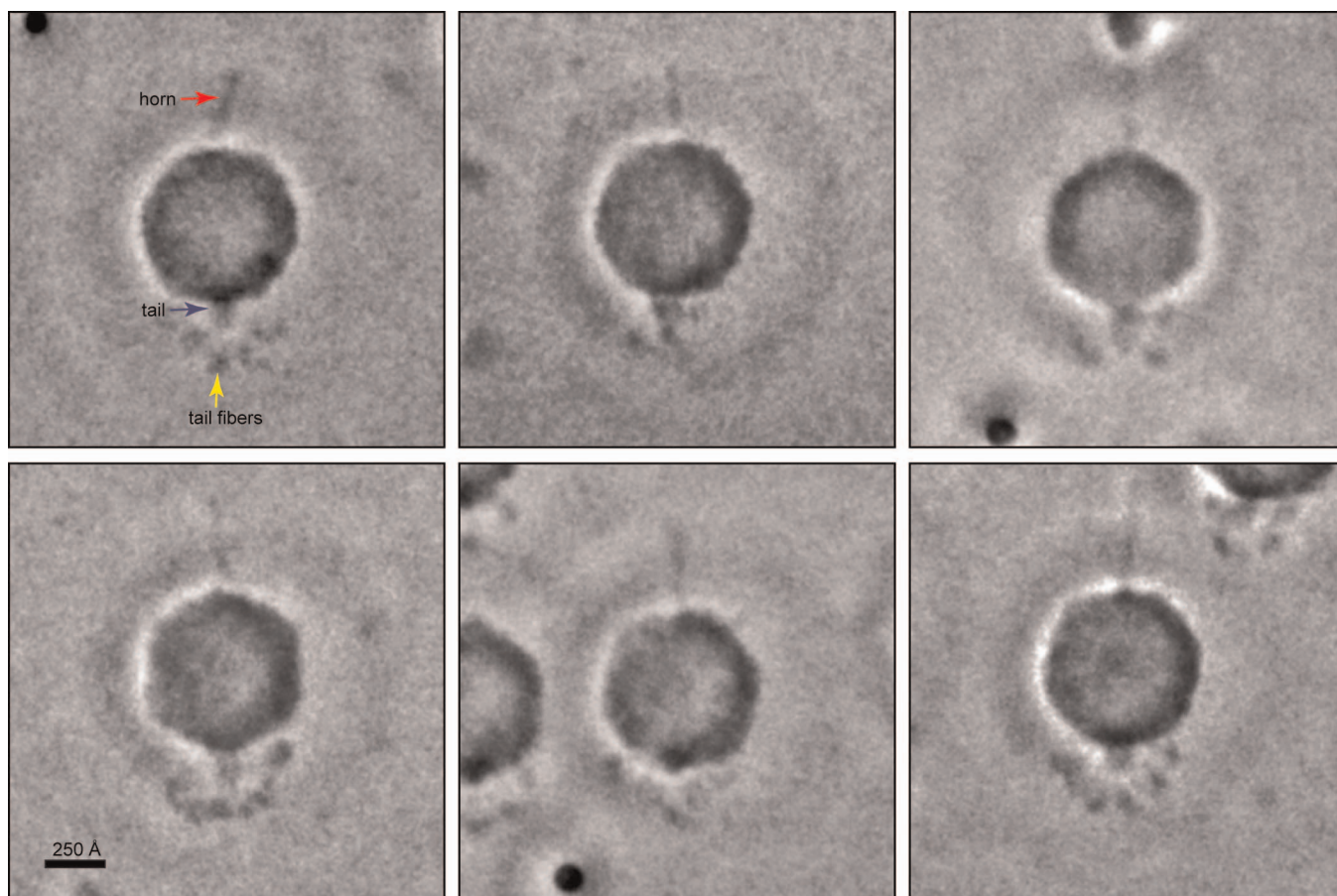
Visualization and segmentation. Visualization of the cell tomograms and averaged maps was done using Chimera³³. Segmentation and annotation of the cell tomograms were done using Avizo (Visualization Sciences Group, FEI).

31. Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 (2012).
32. Baker, M. L. *et al.* Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proc. Natl Acad. Sci. USA* **110**, 12301–12306 (2013).
33. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

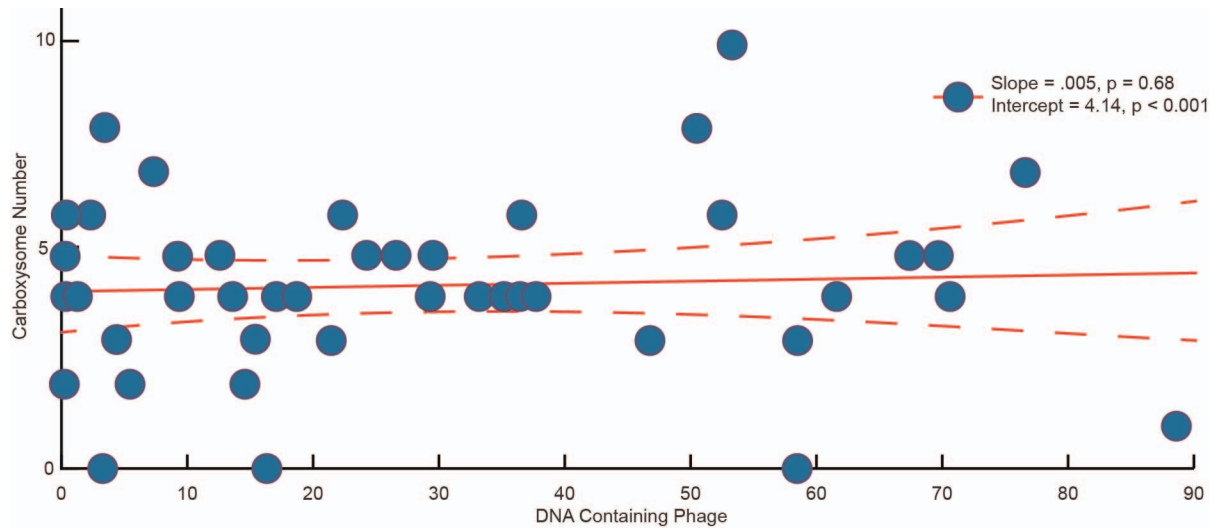


Extended Data Figure 1 | ZPC improves contrast of cryoET images and reveals detailed structural features of Syn5-infected cells. a, A conventional

EM image of a Syn5-infected WH8109 cell. **b,** A ZPC image of the same cell as shown in **a** under the same imaging conditions.



Extended Data Figure 2 | ZPC-cryoEM single-particle images of biochemically purified mature Syn5 phage. The particles are shown with the tail pointing down and the wavy horn pointing up. The tail fibres appear to have variable conformations.



Extended Data Figure 3 | General linear modelling of cellular carboxysome number with progression of infection. The number of carboxysomes remains roughly constant as infection progresses, indicating that their variation does

not correlate with progression of infection. The solid line indicates linear regression; the dashed lines indicate the 95% confidence limits.

CAREERS

TURNING POINT Former English teacher turns to app development **p.713**

GRANTS Peer review fails at predicting success **p.713**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

FOJAB



The MAX IV synchrotron light source in Lund, Sweden, will offer chances for collaboration across the country.

EUROPE

Swedish success story

Institutions shake off rivalries to build scientific collaborations and hire world-class talent.

BY PAUL SMAGLIK

Since the global financial crisis, Sweden has lived in an alternative universe of science funding. While austerity policies have kept research funding levels flat in much of Europe since 2008, Sweden's public science budget has increased by 5 billion Swedish kronor (US\$786 million) over the past 5 years with a rise of another 4 billion kronor to come over the next 5 years. And, as seemingly endless government budget battles have slowed US infrastructure investment, Sweden has seen a building boom. The country has constructed a national high-throughput life-sciences laboratory; begun building new clinical-research laboratories and a hospital; and broken ground on a powerful synchrotron light source and a neutron source.

Now Sweden is increasing international recruitment, backed by public and private money, to fill its facilities and fulfil ambitious research agendas. The Knut and Alice Wallenberg Foundation in Stockholm has been the biggest non-government player in infrastructure investment and international science hiring. Last year, the foundation introduced the Wallenberg Academy Fellows programme to recruit and fund 300 young scientists over 10 years, aiming for 30–50% of the fellows to come from outside Sweden.

The region, however, is adjusting to big changes at pharmaceutical giant AstraZeneca, a long-time presence in Sweden. Since 2010, the company has cut close to 2,000 jobs as it seeks to consolidate all its Swedish research into one facility in Mölndal. But that has provided an incentive for other institutions to take up the

mantle of clinical studies. Former AstraZeneca researchers have translational skills, and Sweden has good databases of individual health records with ample data that are useful in clinical medicine. AstraZeneca's restructuring "gave us an important signal", says Stefan Hansson, vice-dean for medicine at Lund University. "We now maybe need to work more on clinical research. How do we integrate that with our hospitals? How will clinicians add research?"

BOUNCING BACK

Sweden had its own financial struggles in the 1990s, which stymied big growth until the late 2000s. The country must attract world-class scientists from beyond its borders to remain globally competitive, says Göran Sandberg, executive director of the Wallenberg foundation. "We don't have enough bright people ►

► to compete internationally.” Wallenberg and other organizations are aiming to address that shortage through stable funding and elite facilities. “Sweden is far north and it is a small country, but if you give people scientific freedom they are willing to move,” says Sandberg.

David Drew, a New Zealand-born biophysicist, won such freedom with a Wallenberg Academy Fellowship last year. The appointment came with 2.5 million kronor to start his lab at Stockholm University and another 7.5 million kronor to run it for five years, with no need to worry about writing grants, he says. Individual scientists cannot apply for the fellowships: departments of Swedish research institutions make nominations, which are judged by panels that include members of the royal academies. The foundation makes the final decision.

PEOPLE PUSH

The Wallenberg Foundation made its biggest investment in 2003, when it funded the Human Protein Atlas. The project, which has cost the foundation 900 million kronor so far, has characterized 15,000 proteins to date using antibodies, and involves hundreds of scientists.

It served as a test case for large-scale science in Sweden, says Sandberg. The atlas also helped to bring together researchers from institutes including the Karolinska Institute in Stockholm, Stockholm University, Uppsala University and the Swedish Royal Institute of Technology in Stockholm. Together, they formed the Science for Life Laboratory (SciLifeLab).

The network is devoted to high-throughput, interdisciplinary research with translational applications, and has grown to about 1,500 people since its sites in Uppsala and Stockholm were established in 2010. There are still facilities in both cities, but since July SciLifeLab has been managed as a single national laboratory, giving its scientists access to a wide array of expertise and equipment. Director Mathias Uhlén says that the organization is in the process of hiring eight group leaders, and intends to add more scientists next year.

Sweden's relative strength in the global economy should help. “Most ambitious young scientists are going to places where they have the best resources,” says Uhlén. “With the funding we can provide to young researchers, they can thrive.” SciLifeLab receives 340 million kronor a year from the government, a sum that is set to increase to 400 million kronor in 2016. The Wallenberg foundation provides about 120 million kronor yearly, and funding for individual projects can come from other sources.

SciLifeLab awards “very generous” start-up grants, says Sven Nelander, a cancer systems biologist at Uppsala University. This July, he received further support from AstraZeneca, which has established a joint research programme to fund 10 SciLifeLab projects with a total of up to US\$10 million a

year over the next five years.

Scientists joining SciLifeLab will find not only good funds and facilities, but also a collaborative culture, says Uhlén. Stockholm University provides bioinformatics support and research, the Royal Institute of Technology offers high-throughput technical expertise and the Karolinska Institute and Uppsala University both use their clinical bases to ask scientific questions that researchers from all four entities help to answer.

Nelander, for example, is working to develop mathematical models of tumour progression



“You will see more collaboration between technology and health care.”

Stefan Hansson

to shed light on how genetic mutations cause cancer. He uses cell samples from people with cancer at Uppsala University Hospital and takes them to SciLifeLab's Stockholm site, where screening experts help to characterize them. Then he draws on the mathematical expertise of colleagues at the University of Gothenburg, where he was based before he went to Uppsala.

Although Drew is not part of SciLifeLab, he can tap into its expertise to create computer models of the protein-transport process. He can also draw on its high-throughput capabilities to look for small molecules that might bind to transport proteins and inhibit their action.

“We're really building a community,” says Kerstin Lindblad-Toh, co-director of SciLifeLab. “People who work on mathematics in one corner can now work with people who focus on cancer in another.”

INFRASTRUCTURE POTENTIAL

SciLifeLab recruits stand to benefit from other facilities going up in the Stockholm–Uppsala region. The Karolinska Institute is set to open a new university hospital in two years, and construction began last month on a medical-research facility called Biomedicum that will open in 2018 with room for 1,700 researchers.

AstraZeneca maintains a presence in Stockholm. In 2012, the firm established a translational research centre at Karolinska, emphasizing biomarkers. AstraZeneca researchers work alongside Karolinska scientists at the centre, and the company funds several research groups, including graduate students. AstraZeneca is also launching a unit to work on cardiovascular and metabolic disease at Karolinska.

These relationships with academia mark “a new way of operating” for the company, says Anders Ekblom, AstraZeneca's global head of

science and technology integration. He notes, however, that the initiative will not create jobs for people who lost research posts with the firm.

Lund should also see growth in scientific recruitment over the next decade, as two major infrastructure projects come on line. The area already hosts three synchrotron particle accelerators, which use high-speed electrons to create X-ray radiation for spectroscopy and imaging, but in 2010 construction began on a new synchrotron light source: MAX IV, which will open in 2015 and will be the brightest in the world. The aim is to draw researchers from around the world to image materials ranging from proteins to nanoscale molecules.

Lund has also won funding from the European Commission to host the European Spallation Source, a powerful neutron accelerator for materials research that is due to open in 2020. Together, the facilities will create an international hub for materials science, structural biology and nanotechnology.

Hansson says that the European Spallation Source and MAX IV will help to create more of a balance between southwestern Sweden and the more southeasterly Stockholm–Uppsala area. He foresees, for example, researchers from SciLifeLab using the Lund facilities to determine structures of target proteins and molecules that might bind to them. “You will see more collaboration between technology and health care,” Hansson says.

Even with all these vehicles for collaboration, Sweden's science is not free from conflict. Historically, there have been rivalries between institutions in Lund, Stockholm and Uppsala, born mainly of fights over funding and resources. “It's true that there has been some tension,” says Karolinska president Anders Hamsten.

But as more SciLifeLab fellows start to cooperate with sites in other parts of the country, tensions should ease, he says. This summer, SciLifeLab held workshops showing how researchers at other universities can tap into the lab's high throughput resources. Hamsten hopes that Karolinska's involvement in SciLifeLab will help to attract world-class scientists to the institute.

When Drew got word that he had landed the Wallenberg fellowship, he says, he felt gratitude and relief in equal parts: “This was going to be a long-term commitment where the university was serious about me staying.” He remembers thinking, “They want me. It is being matched with a nice level of funding with the vision that this is long term.” The good quality of life in Sweden — including free or inexpensive child-care — was also welcome. And to tackle the long, dark Nordic winters, Drew intends to tap into another benefit: lots of holiday time. He plans to spend a few weeks each Swedish winter in New Zealand. “This way,” says Drew, “I get two summers.” ■

Paul Smaglik is a freelance writer based in Milwaukee, Wisconsin.

CHARLOTTE CARLBERG-BÄRG

TURNING POINT

Mark Matthews

MoodRhythm, a smartphone app that tracks sleep patterns and social routines to improve the mental health of people with bipolar disorder, in June won the US\$100,000 Heritage Open mHealth Challenge for apps that help patients and clinicians to manage chronic disease. Its developer, Mark Matthews, is an information scientist at Cornell University in Ithaca, New York.

How did you end up working on technology to promote mental health?

After I got my English degree at Trinity College Dublin, I taught English in Paris for a year. I saw that it would soon be possible to develop web-based applications that could combine my interest in games and computers with more practical applications. So I went to Dublin City University and did a master of science in multimedia technology and communications. I worked on 'Savant', which used video, audio and three-dimensional (3D) graphics to depict what it is like to have autism. My team won a Europrix Multimedia Award in 2003. That was my first foray into mental health.

How did that lead to your PhD?

I realized that mental-health issues have an incredible stigma, which can make people reluctant to seek help. Technology offers a way for people to maintain privacy, dignity and control over their treatment. What sold me on doing a PhD on human-computer interaction was funding from Ireland's Higher Education Authority and the Vodafone Foundation in Newbury, UK. I developed a computer game to engage teenagers in treatment. Teenagers are often not comfortable talking to strangers. My group and I built a 3D game that allows the teen and therapist to maintain a conversation about personal issues without the need for direct eye-to-eye contact — which was a huge factor in its success. The teenager uses an avatar throughout the therapy-driven game as the therapist looks on.

How do you determine what clinicians need?

It is my job to learn what it is like to be in therapy, to have depression. We involve psychotherapists in our development meetings and I role-play as the patient. I need that to help me understand how to create cutting-edge technologies that will be useful in the clinic.

Was it difficult to secure a research position?

In January 2012 I was at a crossroads. I had applied for a lectureship at University College Dublin, and a Marie Curie Fellowship



at Cornell. I got both in the space of a week. I had to decide between a five-year position in Dublin, where my family is based but the country is experiencing a deep recession, and the opportunity to work in the United States, which proved too good to turn down. Unfortunately that means that I am separated from my partner and young children for long stretches of time.

How did you secure space at Cornell?

The year before I applied for the fellowship, I contacted Geri Gay, the head of the interaction-design lab, who was working on ways to record mood. We Skyped a few times to discuss our mutual interests. Once I got the fellowship, I came to Cornell. It has worked out incredibly well. We are developing ways to use smartphones to detect vocal stress or sleep patterns. One of the great things about coming to the United States is getting to work with people at the top of this field.

How might you build a lab when you return to Dublin?

I know I am not going to walk back into a job, but, in my dream world, I will have a research lab and a commercial venture. I have started thinking about how to get the money together to create a lab. Now that *MoodRhythm* has won the Open mHealth Challenge, I am looking into how to commercialize the software and use the money to branch out. I am also exploring charities and spending time setting up long-term collaborations between researchers at Trinity College, Cornell and the University of Pittsburgh in Pennsylvania. I have no illusions that it will not be difficult, but I will find a way to do high-impact, top-quality research. ■

INTERVIEW BY VIRGINIA GEWIN

SPAIN

Mobility disincentive

Scientists who move between countries or research groups as postdocs take up to nine months longer to achieve tenure at Spanish universities than those who stay put, finds a study (L. Sanz-Menéndez *et al. PLoS ONE* 8, e77028; 2013). The authors surveyed about 1,260 academic scientists and engineers who had earned tenure between 1997 and 2001 at public universities in Spain. Corresponding author Luis Sanz-Menéndez, director of the National Research Council Institute of Public Goods and Policies in Madrid, says that to earn tenure in fewer than five years, the study's average, early-career researchers in Spain should remain at their PhD-granting universities. "There are disincentives for people who are highly mobile," he says, although he notes that moves explicitly directed by a mentor had no negative effects.

GRANTS

Inaccurate predictions

Peer reviewers for biology grant proposals submitted to the US National Science Foundation (NSF) do not accurately predict the apparent success of projects, according to a study (S. M. Scheiner and L. M. Bouchie *Front. Ecol. Environ.* 11, 406–407; 2013). Reviewers' scores and rankings for 41 environmental-biology projects funded by the NSF in 2002 did not correlate well with productivity measures, including the number of publications produced by 2012 and the mean number of citations per year. But reviewers do provide value by weeding out flawed proposals and suggesting improvements, says co-author Samuel Scheiner, programme director for the environmental-biology division of the NSF in Arlington, Virginia.

STUDENTS

Real-time online advice

A platform launched in October offers free real-time personalized counselling to people who want to study in Europe. Users of EduCoach (www.phdportal.eu) can chat to advisers about 2,500 PhD programmes at 950 universities across 40 countries. Topics include programme choice, visas, applications, living arrangements, and language and cultural challenges. EduCoach will launch video chats next year, says Sissy Böttcher, project manager with the platform's European Commission-funded developer, StudyPortals in Eindhoven, the Netherlands.

DEEP IMPRESSIONS

Picture perfect.

BY JOHN GILBEY

I turned the tinted glass tablet over in my hands, looked at the perfectly ground edges, felt the silk-like quality of the cool surface and finally held it up and looked through it. The effect was astonishing, unreal, unexpected.

"I suppose you're going to tell me that this is some sort of quantum shit?"

The suited executive at my side smiled a pained, corporate smile and looked around nervously — as though he expected the chaos of the kitchen to have suddenly sprouted thrusting microphones and whirring news cameras.

"I'm not going to tell you *anything*, John — and I advise you *not* to ask. The terms of this demonstration are *clearly* laid out in the agreement, and we are paying you after all..."

Oh, yes — they were paying all right, a sudden windfall that had landed on my kitchen table like manna from a clear blue sky. Enough money to let me finish the book, buy in some kit for the next project and maybe even fix the chimney.

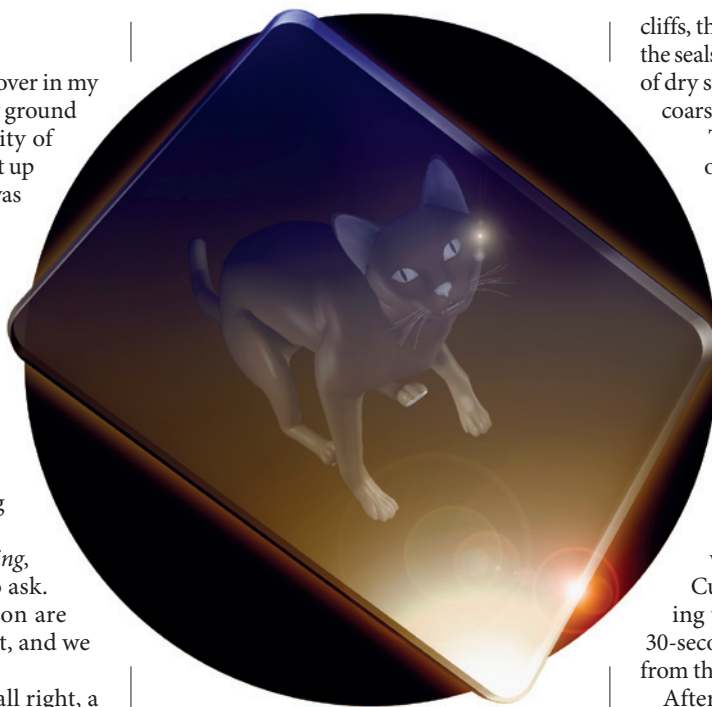
I gently placed the tablet back in its foam-padded case, it made me nervous just holding it. "Does this stuff at least have a name? I've got to call it something in the report."

He gave me a sheepish look. "We're calling it 'OwlPlex' ... Sorry, marketing seemed to think the military would like it."

Yes, I'm sure the military will like it. A sheet of glass, barely a centimetre thick, with no wires or power supply — yet it acted as the highest resolution, most powerful, colour-perfect image intensifier I'd ever seen. And I've seen a few, in a shadowy technical career that I no longer like to think about, let alone talk about. I still had some questions though — whether he wanted them or not.

"Where does the power come from? It can't do this without any input."

He smiled again and pointed out of the window, across the beach and towards the hills. "Sunshine. It soaks up energy when light falls on it, then releases it slowly as the light fades — augmenting the photons that are passing through it at the time. And before you ask, I really can't say any more. Oh, but don't drop it



— it's the only prototype ..." He stood up, made some pointed comments about security and edged his way to the door. "I'll be back next week. Remember, we are looking for insights here. Novel applications, things the kids in Development haven't thought of."

It stayed on the table for four days before I worked up the courage to test it fully. At nightfall I eased the tablet from the case, gripped it carefully in one hand, and turned off the light. One side of the glass glowed like a backlit laptop screen, bright but not too bright, showing me the litter on the floor under the table — including a half chewed mouse that the cat had left there. Making a mental note to clean up, I opened the back door and allowed myself a slow, tuneless whistle as the garden and hillside erupted on the screen into glorious deep autumn hues. The effect was startling, addictive — and I knew exactly what everyone was going to want for Christmas this year.

I awoke in full daylight, with the cat pointedly stropping the end of the sofa in a way that meant he was hungry and wanted feeding *now*. As he ate, I looked around the room wondering what novel ideas I could come up with to justify my fee. The wall next to the fireplace was home to some of my favourite photographs, shot on 5 × 4 film to get the maximum quality. Pictures of the surrounding countryside: the headland with its granite

cliffs, the tranquil sweep of the beach where the seals haul out with their pups, the pattern of dry stone walls enclosing bog cotton and coarse grasses.

The old madness returned. It took only an hour to kludge together a holder for the tablet and secure it like a filter to the pin-sharp prime lens of the ancient technical camera. The fridge still held a box of slow, fine-grained monochrome film, so as darkness fell I loaded up the dark-slides, then hefted the tripod onto one shoulder, picked up the camera and set off.

My assumption was that the tablet would allow a constant exposure, but after the first few photos it was clear that the tablet was fading from a lack of daylight.

Cursing, I increased the shutter timing to compensate — guesstimating a 30-second exposure to harvest a final image from the rapidly dimming tablet.

After development, the images were as sharp as I had hoped — with none of the raster artefacts I would have seen from a conventional system. The content of the last image was, like everything to do with the tablet, unexpected. After a few moments of silent thought I wandered out into the kitchen for tea and a long, long ponder — strongly regretting that I'd given up drinking.

As I looked out over the long-deserted beach in the dawn light, I tried to visualize it as the tablet had imaged it: wooden fishing skiffs hauled out on the shingle, nets drying on the close cropped turf beyond, a single row of whitewashed cottages with split-stone roofs and racks of drying fish. The scrunch of gravel told me that my executive friend was returning, and I was wryly pleased that I'd found his killer application.

I made a bet with myself that no one had taken a time exposure through the fading glow of dying OwlPlex before. Developers today are too hasty, you see — you need to take a long view, just as OwlPlex itself does in those crucial seconds as some weird internal field collapses. In retrospect though, a less distracted man wouldn't have left the tablet so close to the edge of the table. It was all the opportunity the cat needed. ■

John Gilbey writes from the academic seclusion of the University of Rural England. Henry Gee, who provided the inspiration for the story, must share the blame.

➤ **NATURE.COM**
Follow Futures:
@NatureFutures
go.nature.com/mtoodm